In Praise of Bad Codes for Multi-Terminal Communications

Amir Bennatan, Shlomo Shamai (Shitz) and A. Robert Calderbank

***Abstract*— We examine the Gaussian interference channel and the erasure relay channel. We focus on codes that are non-capacity-achieving ("bad") over appropriate point-to-point (two-terminal) channels. Over Gaussian point-to-point channels, for example, such codes require greater SNR than "good" ones to achieve reliable communications, but often exhibit *lower* estimation errors whenever the SNR is below the Shannon limit. Over multi-terminal channels, this advantage of "bad" codes at lower SNRs can be exploited by strategies that apply estimation, at various network nodes, to achieve partial decoding. Such strategies include soft partial interference cancelation (soft-IC) and soft decode-and-forward (soft-DF). We develop variants of these two approaches, which are susceptible to rigorous analysis. We focus on applications of "bad" LDPC codes. We develop analysis tools for soft-DF, including *simultaneous* density evolution (sim-DE), and use standard density evolution to analyze soft-IC. We apply our analysis to the design simple-structured "bad" codes that outperform more complex "good" ones.**

***Index Terms*— Interference channel, relay channel, Gaussian channels, binary erasure channels**

I. Introduction

Multi-terminal wireless networks have attracted great interest in recent years, due to the widespread success of applications like cellular networks, wireless LANs and sensor networks. Research of such networks has drawn heavily from existing results on point-to-point channels, of which our understanding is much more mature. Point-to-point channels are characterized by just two nodes (a source and a destination), that wish to communicate.

In the absence of any additional knowledge, good point-to-point codes would appear to be reasonable candidates for application to multi-terminal channels as well. An examination of the literature, however, reveals that many communication strategies for such channels rely on *bad* codes (see Sections I-A and I-B below). In this paper, we argue that such codes have inherent benefits that often make them better candidates.

In our analysis, we classify codes as "good" if they achieve the capacities of certain point-to-point channels, and "bad" if they do not (a rigorous definition, which involves code-*sequences*, will be provided in Sec. II-D). This choice of terminology, which follows Shamai and Verdú [54], is arbitrary, and intentionally ignores many other attributes of the codes, like availability of low-complexity decoding algorithms (our use of quotes in "good" and "bad" reflects this fact). Throughout the paper, the context of our classification of codes as "good" or "bad" is always point-to-point channels, even when the subject of the discussion is multi-terminal channels.

We begin below with a description of the problem, followed by our motivation for addressing it in Sec. I-B. Our main contributions in this paper are summarized in Sec. I-C.

A. Background and Problem Formulation

To illustrate some of the benefits of "bad" codes in multi-terminal scenarios, we begin by considering a *point-to-point* scenario. We focus on the minimum mean-square error (MMSE) of an estimate, computed at the destination, of the codeword transmitted from the source. That is, we assume the destination undertakes a less ambitious task than frequently found in information-theoretic literature: It attempts to estimate the transmitted codeword as best it can, rather than decode it completely. When reliable decoding is possible, the MMSE will of course be close to zero.

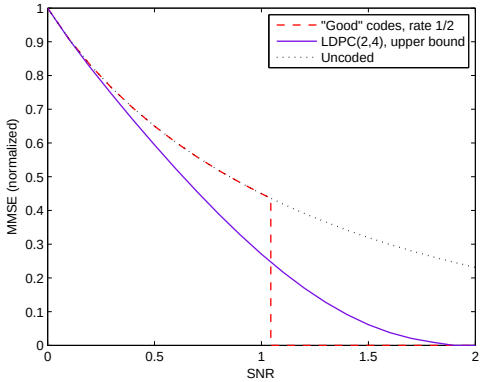


Fig. 1. MMSE as a function of SNR in a point-to-point BIAWGN channel. Details are provided in Appendix I.

In the discussion below, we assume that the channel is a binary-input additive white Gaussian noise (BIAWGN) channel (later, we will examine an additional model). The MMSE is clearly a function of the SNR over this channel. Fig. 1 plots the MMSE curve in three different scenarios (see Appendix I for rigorous details). The first curve corresponds to

The work of S. Shamai (Shitz) is supported by the Israel Science Foundation (ISF) and NEWCOM++ EU 7th Framework Program. The work of R. Calderbank is supported in part by NSF under grant DMS 0701226, by ONR under grant N00173-06-1-G006, and by AFOSR under grant FA9550-05-1-0443.

A. Bennatan and R. Calderbank were with the Program in Applied and Computational Mathematics (PACM), Princeton University, Princeton, NJ 08540 USA. A. Bennatan is now with Samsung Israel R&D Center (SIRC), Ramat Gan, 52522, Israel (e-mail: amir.b@samsung.com). R. Calderbank is with the Department of Computer Science, Duke University, Durham, NC 27708 USA (e-mail: robert.calderbank@duke.edu). S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

a rate-$1/2$ "good" code[1]. Surprisingly, it doesn't matter *which* "good" code. Peleg $et\ al.$ [45] have shown that the optimal estimation error, with *any* "good" code can be determined entirely from its rate. Their proof follows from the relation between the mutual information and the MMSE, as established by Guo $et\ al.$ [23] and also observed by Measson $et\ al.$ [39]. The second curve is an upper bound on the MMSE of a rate-$1/2$ LDPC (2,4) code, which is known to be very "bad"[2]. The third curve corresponds to transmission of a stream of uncoded bipolar (BPSK) bits.

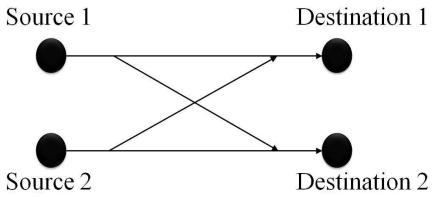
As expected, the curve corresponding to uncoded communications, upper bounds the two other curves. With such communications, the destination does not have the benefit of a code structure to draw upon in its computations. At SNRs above the Shannon limit (SNR > 1.044), the rate-1/2 "good" code's MMSE (which is zero) outperforms the LDPC upper bound. This coincides with the intuition that "good" codes are better than "bad" ones. However, an interesting phenomenon occurs at low SNRs. While the "good" code's MMSE provably collapses to that of uncoded communications, the "bad" LDPC code exhibits graceful degradation and achieves a substantially better MMSE.

A similar observation was made by Berrou $et\ al.$ [5, Sec. II.B] in their choice of the components of turbo codes, for point-to-point communications. The "bad" codes they used, however, were combined and manipulated (by parallel concatenation) to produce other, relatively "good" codes. When "bad" codes are used unaltered, their above-mentioned advantage is generally meaningless in point-to-point communications. In such scenarios, we are typically interested in complete, reliable decoding[3] (an asymptotically zero error probability), and so all non-zero levels of MMSE are equally unacceptable.

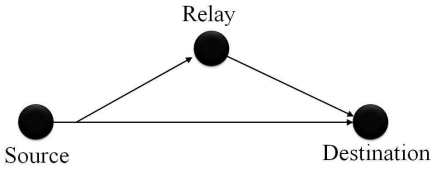
With multi-terminal communications, however, estimates may be perceived as a form of *partial*-decoding, and computed at nodes other than the destinations of a given message, as byproducts of communication strategies. In this context, the advantage of "bad" codes may be meaningful. In this paper we are interested in quantifying the advantage in terms of achievable rates at destination nodes.

In our analysis, we focus on two simple multi-terminal channel models: The interference channel, as introduced by Shannon [55], and the relay channel, as introduced by Van Der Meulen [58]. Both channels are illustrated in Fig. 2.

An interference channel is characterized by two pairs of nodes, each pair consisting of a source and destination that wish to communicate. Unlike point-to-point channels, each destination experiences interference resulting from the signal produced by the source of the other pair. In a relay channel, a single pair of source and destination nodes wish to communicate, but are aided by a *relay* node which lends its resources to support their communications. These channels capture two of the fundamental phenomena that characterize wireless networks: Interference between nodes (e.g. resulting from the shared wireless medium) and the potential of cooperation to achieve better performance. The capacities of both channels are in general still unknown.



(a) An interference channel.



(b) A relay channel.

Fig. 2. Two multi-terminal channel models.

In this paper, we focus on two classes of communication strategies. Soft interference cancelation (soft-IC) for interference channels, and soft decode-and-forward (soft-DF) for relay channels. With soft-IC, each of the destinations in an interference channel computes a soft estimate of the interfering codeword as a byproduct of its decoding algorithm. With soft-DF, the relay uses its observed signal to compute a soft estimate of the signal transmitted by the source. Thus, both strategies use estimation as a method for partial decoding.

While an overwhelming body of research exists on soft-IC, the majority of it (e.g. [61], [6], [10], [29], [52]) focuses on the approach's application as a component of larger, iterative schemes designed to achieve complete (rather than partial) decoding of multiple signals. The concepts of soft-IC for the purpose of partial decoding of interference, as considered in this paper, were proposed by Divsalar $et\ al.$ [15] as well as [34], [62], [24]. Soft-DF was proposed by Sneessens and Vandendorpe [56], and related concepts were also examined in [60], [65], [41], [35][4].

Partial decoding, as applied by soft-IC and soft-DF, is a useful compromise in cases where complete decoding would be desirable if possible, but is not required by the terms of the problem. In an interference channel, decoding of the interfering signal at each destination, would enable the node to eliminate its interference. In a relay channel, decoding of the signal transmitted by the source, would enable the relay to better assist the source in the delivery of the associated message to the destination. Both signals, however, are not required at the respective nodes, and may be discarded once communication is over. Insisting on their complete decoding often imposes a burden on the communication strategy, which may outweigh the signals' usefulness. Partial decoding is a

[1]To simplify our discussion in this section, we neglect some rigorous details. A precise description of the curve is provided in Appendix I and involves the asymptotic normalized MMSE of a "good" code-*sequence*.

[2]This follows from the analysis of [9], [53].

[3]One exception is joint source-channel coding, where "bad" codes were considered [28].

[4]Similar concepts were also examined by Gomadam and Jafar [22]. In their work, however, estimation is performed individually on each received symbol, and dependence between symbols implied by the code is not exploited.

means of balancing these two considerations.

Our interest in partial decoding also follows from its central role in other communication strategies, beside soft-IC and soft-DF. The best-known rates for the interference and relay channels are achieved by Han-Kobayashi (HK) [25] and partial decode-and-forward[5] (partial-DF) [12, Theorem 7],[31] respectively. Both strategies, however, achieve partial decoding by *rate-splitting*[6], rather than estimation as discussed above. Rate-splitting involves using codes that were each constructed by combining two (generally, two or more) other, auxiliary codes. With HK, for example, this gives each destination the option of decoding *one* of the two auxiliary codewords that constitute the interfering signal (in addition to the two that constitute the signal from its own source), amounting to a partial decoding of the interference. From a practical perspective, however, soft-IC and soft-DF, enjoy a number of advantages over HK and partial-DF which will be discussed in Sec. I-B below.

Interestingly, many implementations of soft-DF (e.g. [56]) involve low constraint length convolutional codes (constraint-4 in [56]), which as discussed in Sec. I-B below, are point-to-point "bad". In [56], [60], [41], [35] and [15], [34], [62], [24], efficient practical implementations of soft-DF and soft-IC were proposed, and extensive simulations were reported that confirm the methods' effectiveness.

A rigorous analysis of these strategies, however, is often difficult to achieve[7] (see Sec. I-C for a discussion of some of the difficulties). In this paper, we nonetheless develop rigorously-proven bounds on the performance of soft-DF and soft-IC, in terms of achievable communication rates at asymptotically large block lengths. Like [56] we focus on applications of soft-DF (as well as soft-IC) that involve "bad" codes. Furthermore, we place a specific emphasis on a comparison with "good" codes.

B. Motivation

Our interest in this problem was motivated in part by the potential for achieving a better tradeoff between achievable rates and decoding complexity, in comparison to point-to-point scenarios.

With many classes of codes, "goodness" of point-to-point performance and decoding complexity, are both related to the complexity of the codes' structures. The constraint length of a convolutional code [59], for example, is a measure of the complexity of its structure. Codes with higher constraint lengths have complex trellis diagrams, and thus more complex structures. Similarly, the density of an LDPC code's parity check matrix [48], as measured by the average weight (number of ones) in each row, is arguably a measure of the complexity of its structure (higher density meaning greater complexity). Over point-to-point channels, theoretical results for both classes of codes, imply that the complexities of their structures must approach infinity as their code rates approach capacity, for reliable communication to be possible [20, Theorem 3.3],[9],[53], [59, Sec. 5.4]. However, the decoding complexities of both LDPC and convolutional codes (via the belief-propagation and Viterbi algorithms, respectively) grow unboundedly with the complexities of their structures[8] (as well as with their block lengths) [20][19].

Thus, from the perspective of decoding complexity, *simple*-structured codes are advantageous. Formally, we define a *sequence* of codes to be *simple*-structured if the complexities of its codes are bounded, and *complex*-structured if they are not. By our above discussion, simple-structured codes are bounded away from the capacities of point-to-point channels, i.e. they are "bad".

Over multi-terminal channels, however, our discussion in Sec. I-A implies that simple-structured "bad" codes sometimes exhibit advantages in terms of partial-decoding at various network nodes, which could perhaps be used to compensate for their weaknesses. In our analysis of soft-IC and soft-DF in this paper, we provide examples of simple-structured LDPC codes whose performance, in terms of achievable communication rates, provably surpasses complex-structured "good" codes.

Our results raise the possibility that simple-structured codes fare better over multi-terminal channels, in terms of the gap from capacity, than they do over point-to-point channels. If so, their simple structures may be applied to improve the tradeoff between achievable rates and decoding complexity in multi-terminal scenarios. In this paper, we do *not* resolve these questions. Our objective is to motivate further research into the problem, and into benefits of simple-structured codes.

Note that rate-splitting, as used by HK [25] and partial-DF [12, Theorem 7], also frequently yields "bad" codes. In an example provided in Sec. V-B, the best application of HK that we found, relies on provably "bad" codes. The applications of rate-splitting in [25] and [12, Theorem 7], however, involve randomly-generated auxiliary codes. Unlike the simple-structured "bad" codes discussed above, such codes are unstructured, and no low-complexity decoding algorithms are known for them. Thus, at least with respect to these applications of HK and partial-DF, soft-IC and soft-DF as described e.g. in [56] and [62], have an advantage.

Soft-IC enjoys an additional advantage over HK. With HK, partial decoding implies that each receiver must jointly decode three codewords, two from its own source and one that constitutes a part of the interference. Soft-IC requires it to examine just *two* codewords; decoding one, and estimating the other (the interference). As the computation time of many algorithms for decoding and estimation (e.g. [6]) grows exponentially with the number of codewords jointly examined, this constitutes an advantage[9].

Beyond our interest in decoding complexity, our analysis may shed light on the best achievable rates (capacities)

[5]In [31] it is known as *multipath* decode-and-forward.

[6]This term was coined by Rimoldi and Urbanke [50], in the context of coding for multiple-access channels.

[7]Note that in [65], simulation results were augmented by an EXIT chart analysis. EXIT charts, however, rely on heuristic assumptions and do not constitute rigorous analysis.

[8]Despite the remarkable performance of many LDPC codes (e.g. [49]) at rates that are very close to capacity, their densities imply impossible decoding complexity.

[9]Note that joint decoding can be avoided by resorting to further rate-splitting, using concepts similar to the ones suggested by [50]. However, such methods are suboptimal at all but asymptotically large block lengths.

of multi-terminal channels. As mentioned above, the best rigorously-proven known results for the interference and relay channels, were obtained by applications of HK and partial-DF (respectively) that involve randomly generated codes. Recently, Philosof and Zamir [46], Nazer and Gastpar [42] and Narayanan $et\ al.$ [40] demonstrated that structured codes sometimes exhibit advantages over random ones, in various communication scenarios. Our focus on *simple*-structured codes is different from theirs. However, as the capacities of both the interference and relay channels are yet unknown, an analysis which builds on our methods may be of similar theoretical interest.

C. Overview of Main Results

Our focus in this paper is thus on the performance of soft-DF and soft-IC, in terms of achievable communication rates, when used with simple-structured "bad" codes. An analysis of soft-DF is complicated by a number of factors. First, standard information-theoretic techniques do not straightforwardly apply to non-random simple-structured codes.

More importantly, however, analysis of soft-DF is complicated by unfavorable attributes of soft decoding. Consider the relay's estimation of the codeword transmitted by the source. As explained later, this estimate is conveyed to the destination, which relies on it in its decoding process. Ideally, we would treat the estimation error as additive white noise, and apply standard coding-theoretic analysis techniques. However, the estimation error is very different from such noise. The components of the error vector can generally be shown to be strongly correlated[10], and the correlation patterns are complex. The error is not independent of the transmitted codeword. Lastly, the error cannot be argued to be independent of the codebook in use, because the estimation process relies on the structure of the code. In this paper we develop methods for overcoming these difficulties.

To enable rigorous analysis, we develop a variation of soft-DF that is analytically tractable. Our variation assumes the use of LDPC codes [20], and relies on their associated belief-propagation (BP) algorithm as a method of soft estimation. We thus refer to it as soft-DF-BP. Our analysis of this algorithm applies to erasure relay channels (defined in Sec. III-A below).

Our choice to focus on LDPC codes follows from the elaborate design and analysis tools that exist for them (e.g. [47]). This choice may appear unusual, as LDPC codes are known primarily for their relative "goodness", i.e. the possibility of capacity-approaching performance over many point-to-point channels (e.g. [49],[36]). However, as noted in Sec. I-B, "bad" LDPC codes exist as well. Such codes can be designed by manipulating the *edge distributions* that determine the structure of their underlying Tanner graphs (see Sec. II-E).

BP in the literature is typically used for complete decoding, not for soft estimation. However, the algorithm in fact approximates bitwise maximum *a-posteriori* (MAP) decisions, and so is actually an estimation algorithm. Over point-to-point channels, BP has typically been applied in scenarios where the level of noise is low enough for the proportion (fraction) of erroneously decoded bits to be small, essential amounting to complete decoding. In this paper, we will examine its performance in other scenarios as well[11].

Our definition of soft-DF-BP is based on compress-and-forward (CF) [12], [32], [26]. With CF (see Sec. III-A), the relay forwards its channel output vector to the destination. The destination combines this with its own channel observation, and attempts to decode using both. With soft-DF-BP (see Sec. III-B), the relay first applies BP to estimate the transmitted codeword. While complete decoding is in general not possible, estimation reduces of the level of noise, thus improving the quality of the signal delivered to the destination.

Our analysis of soft-DF-BP focuses on the performance of *simultaneous BP* (sim-BP), a hypothetical algorithm which accesses the channel outputs from both the relay and the destination. In practice, the two nodes are physically separated, and so the algorithm cannot be realized. It is nonetheless designed so that its performance can be used to upper bound the number of bit errors at the output of soft-DF-BP. Furthermore, the structure of sim-BP enables its analysis using an extension of density evolution, a method devised for the analysis of LDPC codes over point-to-point channels [47]. We refer to the extension as *simultaneous density evolution* (sim-DE).

As with CF, with soft-DF-BP the relay exploits its channel to the destination, to deliver the estimate it computed to that node. To reduce the demands on bandwidth (rate) to fit the available capacity of this channel, the delivered signal is first compressed and distorted (lossy compression). In our analysis, we demonstrate that the simple structure of "bad" LDPC codes can often be applied to improve the compression rate and reduce the distortion.

Turning to soft-IC, its analysis is simplified by powerful analysis tools that were developed in the context of multiuser detection, e.g. by Boutros and Caire [6] and Amraoui $et\ al.$ [1] (and references therein). We focus on a variation of soft-IC which essentially coincides with joint iterative multi-user detection (iterative-MUD) as suggested in these references. In our context, we refer to the algorithm as soft-IC-BP. This choice of terminology reflects the algorithm's role in our setting, as explained below. Our analysis applies to symmetric BIAWGN interference channels.

Iterative-MUD is typically applied in multiple-access settings, where a destination attempts to decode (completely) a superposition of two (or more) signals from different users. Like BP, however, iterative-MUD in fact approximates bitwise MAP decisions, and thus straightforwardly applies to estimation as well. With soft-IC-BP, each of the two destinations applies the algorithm to attempt to decode the codewords from both sources. Unlike multiple-access settings, with soft-IC-BP we relax the requirement of complete decoding of the interfering codeword, and tolerate a large error in its estimation. Analysis tools of iterative-MUD carry over straightforwardly to the analysis soft-IC-BP.

As benchmarks for comparison, we consider strategies that

[10]This follows because optimal estimation is in general not achieved by symbol-wise computation.

[11]A similar approach was taken by Barak $et\ al.$ [3], in the context of communication over erasure channels with unknown erasure probabilities.

rely on "good" codes. Over symmetric BIAWGN interference channels, we examine single user detection (SUD) and multiuser detection (MUD)[12] (see Sec. IV-A). Furthermore, we develop tight bounds on the achievable rates of communication with *any* set of "good" codes, and prove that they cannot exceed the performance of SUD and MUD. Over erasure relay channels, we compare our performance with the above-mentioned CF, as well as with decode-and-forward (DF) (see Sec. III-A). We also provide bounds on applications of soft-DF-BP that use "good" codes, which are valid under certain plausible assumptions.

To demonstrate the effectiveness of our constructions, we design specific applications of soft-DF-BP and soft-IC-BP, that provably outperform the above benchmarks. As with communication over point-to-point channels, the identities of the LDPC codes in use play a crucial role in the performance of soft-DF-BP and soft-IC-BP. To design effective codes, we extend a technique proposed by Richardson $et\ al.$ [49] from point-to-point channels to our settings.

As noted in Sec. I-B, our objective in this paper is to demonstrate the potential of simple-structured codes, for which we believe low-complexity algorithms exist. Design of such algorithms, however, is beyond the scope of this work. Specifically, with soft-DF-BP, our discussion leaves out the details of compression at the relay and communication of the estimate to the destination. We assume that these components of the strategy are achieved in the same way as CF, which applies high-complexity computations.

This paper is organized as follows. In Sec. II we introduce some preliminary notations and definitions. We formally define "good" and "bad" codes, and provide some relevant background on LDPC codes. In Sec. III we define soft-DF-BP, develop its analysis tools, and derive our bounds on the performance of "good" codes. In Sec. IV we do the same with soft-IC-BP. In Sec. V we present specific applications of soft-DF-BP and soft-IC-BP. Finally, Sec. VI concludes the paper. Throughout the paper, proofs and various details are deferred to the appendix.

II. Preliminaries

A. General Notation

Vector values are denoted by boldface (e.g. $\mathbf{x}$) and scalars by normalface (e.g. x). Random variables are upper-cased (X) and their instantiations lower-cased (x). $\mathbb{E}$ denoted expectation. $\exp(x)$ denotes the exponential function, e^x (we will use both notations interchangeably). $\ln$ denotes the natural logarithm (to the base e) and $\log$ denotes the base 2 logarithm. Correspondingly, all communication rates are given in bits per channel use. $[a,b]$ denotes the interval $\{x \in \mathbb{R} : a \leq x \leq b\}$, and (a,b) denotes $\{x \in \mathbb{R} : a < x < b\}$.

Given a node i in a graph, $\mathcal{N}(i)$ is the set of nodes that are adjacent to i. Given a vector $\mathbf{x} = (x_1, ..., x_n)$, we let the vector $\mathbf{x}_{\sim i}$ denote the vector obtained from $\mathbf{x}$ by omitting x_i, that is,

$$\mathbf{x}_{\sim i} \triangleq (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) \tag{1}$$

[12]While HK often also involves "good" codes, they are combined by rate-splitting to produce other, typically "bad", codes.

$h(x)$ denotes the binary entropy function,

$$h(x) = -x \cdot \log x - (1-x)\log(1-x)$$

n will typically denote the block length of a code, whose identity will be clear from the context. $o(1)$ is a term that approaches zero with n.

B. Binary Input AWGN and Binary Erasure Channels

We now define two point to point channels which we will use throughout the paper to classify codes as "good" or "bad". The binary additive white Gaussian noise (BIAWGN) channel is characterized by the equation,

$$Y = X + Z \tag{2}$$

where Y is the channel output, X (the transmitted signal) is taken from $\{\pm 1\}$, and Z is a zero-mean real-valued Gaussian random variable with variance σ^2, whose realizations at different time instances are statistically independent. σ is a positive constant.

The binary erasure channel (BEC) is characterized by,

$$Y = \begin{cases} e, & \text{with probability } \delta \\ X, & \text{otherwise.} \end{cases} \tag{3}$$

where Y is the channel output. X is the channel input, and is taken from $\{0,1\}$. $\delta \in [0,1]$ is a constant. e is a symbol indicating an "erasure" event. We assume that the channel transitions at different time instances are independent. We let BEC(δ) denote a BEC with crossover probability δ.

We define the Shannon limit for the BIAWGNC and BEC in the usual way, as the inverse of the Shannon capacity function:

Definition 1: Let $R \in [0,1]$. The BIAWGN (resp. BEC) *Shannon limit* for rate R is the minimal (resp. maximal) value SNR* (resp. erasure probability $\delta^\star$) such that reliable communication is possible at rate R.

C. Notations for Analysis of Erasures

The following notations will be useful in our analysis of erasure channels. For simplicity, we rewrite (3) as,

$$Y = X + E \tag{4}$$

where E is an *erasure noise* random variable, denoted $E \sim$ Erasure(δ) and equal to e with probability δ and to 0 otherwise, and addition of two values $x_1, x_2 \in \{0,1,e\}$ is defined as,

$$x_1 + x_2 \quad \triangleq \quad \begin{cases} e, & x_1 = e \text{ or } x_2 = e \\ x_1 \oplus x_2, & \text{otherwise.} \end{cases}$$

where $\oplus$ denotes modulo-2 addition. Note that the sum of two independent erasure noise variables $\mathrm{E}_1 \sim$ Erasure(δ_1) and $\mathrm{E}_2 \sim$ Erasure(δ_2) is also an erasure noise, distributed as Erasure($\delta_1 \circ \delta_2$) where,

$$\delta_1 \circ \delta_2 \quad \triangleq \quad \delta_1 + \delta_2 \cdot (1 - \delta_1) \tag{5}$$

We also define the multiplication of two values $x_1, x_2 \in \{0, 1, e\}$ as follows,

$$x_1 \cdot x_2 \quad \triangleq \quad \begin{cases} x_1 \cdot x_2, & x_1 \neq e \text{ and } x_2 \neq e \\ x_2, & x_1 = e \\ x_1, & x_2 = e \end{cases} \tag{6}$$

Note that although the product $x_1 \cdot x_2$ is defined also for cases that x_1 and x_2 are not erasures and $x_1 \neq x_2$, we will not encounter such cases in practice. We define the product between two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \{0, 1, e\}^n$ as the vector obtained by multiplying the respective components.

Finally, we introduce two more definitions,

Definition 2: Let $\mathbf{x} \in \{0, 1, e\}^n$. The *erasure rate* of $\mathbf{x}$, denoted $P_e(\mathbf{x})$, is the fraction of its components that are equal to an erasure.

Definition 3:

1) Let $x, y \in \{0, 1, e\}$. We say that y is *degraded* with respect to x if the following two conditions do not hold simultaneously: $x = e$ and $y \neq e$.
2) If $\mathbf{x} \in \{0, 1, e\}^n$ and $\mathbf{y} \in \{0, 1, e\}^n$, we say that $\mathbf{y}$ is *degraded* with respect to $\mathbf{x}$ if for all $i = 1, ..., n$, y_i is degraded with respect to x_i. Equivalently, $\mathbf{y}$ is degraded with respect to $\mathbf{x}$ if the set of indices of $\mathbf{y}$ that are erasures, contains the equivalent set for $\mathbf{x}$.

D. "Good" Codes

Our definition of "good" codes is a variation of the definition of Shamai and Verdú [54]. For simplicity, we have specialized it for BIAWGNCs and BECs, but it can straightforwardly be generalized to other classes of channels as well. The definition focuses on *sequences* of codes $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$. Given such a sequence, we define its *rate* as the limit of the rates of the individual codes, if the limit exists.

Definition 4: Let $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$ be a sequence of codes of rate R.

1) We say that $\mathcal{C}$ is *"good" for the BIAWGN channel* if the following holds:
$$\lim_{n\to\infty} P_e(\mathcal{C}_n; \delta) = 0, \qquad \forall \text{SNR} \geq \text{SNR}^\star$$
where $\text{SNR}^\star$ is the Shannon limit for the BIAWGN at rate R and $P_e(\mathcal{C}_n; \text{SNR})$ is the probability of error under maximum-likelihood (ML) decoding, when the code $\mathcal{C}_n$ is used over a BIAWGN with the specified SNR.
2) We say that $\mathcal{C}$ is *"good" for the BEC* if the following holds:
$$\lim_{n\to\infty} P_e(\mathcal{C}_n; \delta) = 0, \qquad \forall \delta \leq \delta^\star$$
where $\delta^\star = 1 - R$ is the Shannon limit for the BEC of rate R and $P_e(\mathcal{C}_n; \delta)$ is the probability of error under ML decoding, when the code $\mathcal{C}_n$ is used over a BEC with an erasure probability of δ.

Remark 1: For simplicity of notation, we adopt the convention that the block length of $\mathcal{C}_n$ is n.

We refer to a code-sequence as "bad" for a particular class of channels if it is not "good". We will occasionally use the terms "good" or "bad" without mentioning the class of channel, whenever the class will be clear from the context. Specifically, when discussing erasure relay channels, "goodness" will be assumed to relate to the BEC, and when discussing BIAWGN interference channels, "goodness" will relate to the BIAWGN.

The existence of "good" code-sequences is guaranteed by the achievability proof of channel capacity (e.g. [13])[13]. Furthermore, BIAWGN and BEC channels admit "good" sequences of *linear* codes[14].

E. LDPC Codes

LDPC codes play a central role in our analysis. A comprehensive review of these codes is available e.g. [48]. For completeness, we now describe some of their essential features, which we will use in our analysis.

An LDPC code is characterized by a bipartite *Tanner graph* [57], as in Fig. 3. The nodes on its left side are called *variable nodes*, and each corresponds to a transmitted codebit. The nodes on the right are *check nodes*, and each corresponds to a parity-check. The codewords of the LDPC code are defined by the condition that at each check node, the set of codebits corresponding to adjacent variable nodes, must sum to zero (modulo-2).

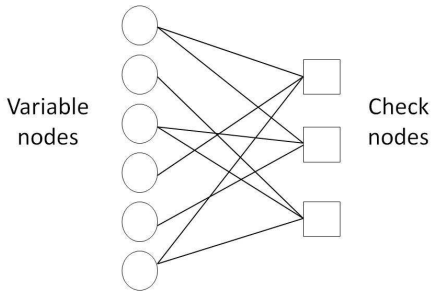


Fig. 3. An example of the Tanner graph of an LDPC code.

The performance of an LDPC code is determined by the structure of its Tanner graph. Luby *et al.* [37] suggested graphs characterized by two probability vectors, $\lambda = (\lambda_1, \ldots, \lambda_c)$ and $\rho = (\rho_1, \ldots, \rho_d)$, which are known as *edge distributions*. In a (λ, ρ) Tanner graph, for each $i = 1, ..., c$ a fraction λ_i of the edges has left degree i, meaning that they are connected to a variable node of degree i. Similarly, for each $j = 1, ..., d$ a fraction ρ_j of the edges has right degree j, meaning that they are connected to check node of degree j. While λ_i refers to the fraction of edges, the fraction of variable nodes of degree i is in general different, and can be shown to equal,

$$\tilde{\lambda}_i = \frac{\lambda_i / i}{\sum_{k=1}^{c} (\lambda_k / k)}. \tag{7}$$

[13]To make this statement precise by our above definition of "good" codes, we invoke the well-known fact that the capacity-achieving distribution for all BIAWGN and erasure channels is the same (Bernoulli(1/2)) [21, Theorem 4.5.1].

[14]This was proven by Elias [16] for binary symmetric channels and later extended to other binary-input symmetric-output channels (see e.g. [59]).

The fraction $\tilde{\rho}_j$ of check nodes of degree j can similarly be obtained from ρ.

A Tanner graph is said to be (c,d)-*regular* if all the variable nodes have degree c and all the check nodes have degree d. Equivalently, the graph is characterized by a pair (λ,ρ) where $\lambda_c = 1$ and $\rho_d = 1$. A code is said to be (λ,d)-*right-regular* if it is characterized by (λ,ρ) where $\rho_d = 1$, i.e. all check-nodes have degree d.

The LDPC (λ,ρ) *ensemble* is the set of codes whose Tanner graphs are characterized by (λ,ρ). As often encountered in information theory, analysis of LDPC codes is greatly simplified by focusing on the average performance of a code selected at random from such an ensemble, rather than on the performance of an individual code. Luby $et\ al.$ [37, Sec. III.A] suggested a procedure for randomly generating a Tanner graph that is characterized by a given (λ,ρ). Different pairs (λ,ρ) may correspond to substantially different performance, and so much of the analysis of LDPC codes focuses on finding effective pairs.

The rate of a (λ,ρ) LDPC code can be shown to be lower bounded by the following value, known as the *design rate*.

$$R_{\text{design}} = 1 - \frac{\sum_j \rho_j/j}{\sum_i \lambda_i/i} \tag{8}$$

Central to the success of LDPC code has been their efficient belief-propagation (BP) decoding algorithm. A general description of the algorithm is available e.g. [8, Algorithm 2]. In the special case of transmission over the BEC, the algorithm has a simple equivalent formulation [37], provided below. The input to the algorithm is the channel output vector $\mathbf{y} = [y_1,...,y_n]$ and the output is a vector $\mathbf{y}^{\text{BP}}$ of decisions (estimates) for the various bits. In the description below we make use of notation which was introduced in Sec. II-C.

Algorithm 1 (Belief-propagation (BP) over the BEC):

1) **Iterations.** Perform the following steps, alternately, a pre-determined t times.
   - *Rightbound iteration number* $\ell = 0,...,t-1$. At all edges (i,j) compute the rightbound messages $\mathrm{r}_{ij}^{(\ell)}$ as follows,

$$\mathrm{r}_{ij}^{(\ell)} = \begin{cases} y_i, & \ell = 0, \\ y_i \cdot \prod_{j' \in \mathcal{N}(i)\setminus\{j\}} \mathrm{l}_{j'i}^{(\ell)}, & \ell > 0. \end{cases} \tag{9}$$

     where $\mathrm{l}_{j'i}^{(\ell)}$ is a leftbound message computed in the preceding leftbound iteration.
   - *Leftbound iteration number* $\ell = 1,...,t$. At all edges (j,i) compute leftbound messages $\mathrm{l}_{ji}^{(\ell)}$ as follows,

$$\mathrm{l}_{ji}^{(\ell)} = \sum_{i' \in \mathcal{N}(j)\setminus\{i\}} \mathrm{r}_{i'j}^{(\ell-1)} \tag{10}$$

2) **Final decisions.** For each $i = 1,...,n$ compute,

$$y_i^{\text{BP}} = y_i \cdot \prod_{j \in \mathcal{N}(i)} \mathrm{l}_{ji}^{(t)} \tag{11}$$

Note that the right hand side of (11) may potentially be an erasure. In some formulations of the BP algorithm, y_i^{BP} is randomly set to 0 or 1 whenever this happens. In this paper, however, we allow y_i^{BP} to remain an erasure. As noted in Sec. I-C, our analysis will include cases where many of the components of $\mathbf{y}^{\text{BP}}$ remain erasures, amounting to an incomplete decoding of the transmitted codeword.

III. Coding for the Erasure Relay Channel

A. Channel Model and Achievable Strategies

Fig. 4 depicts our model for the binary erasure relay channel. It is a variation of the model suggested by Kramer [30] and is a special case of the models of [27], [12]. The channel is characterized by a triplet (δ_2,δ_3,C_o), explained below.

Like [27], we assume that the channels to the destination from the source and from the relay are decoupled. That is, the destination receives two independent channel observations: Y_3, which is a function of the source signal X_1 and Y_3' which is a function of the relay signal X_2. As we will see, this assumption simplifies the analysis, while retaining the essential challenges facing the design of relay communication strategies. Following [27], we characterize the channel from the relay to the destination by its capacity C_o alone, and refrain from specifying the channel transition probabilities. The precise probabilities will be inconsequential, and our strategies will apply equally regardless of them (as [27]).

We assume that the source signal X_1 is received by the relay and the destination via independent memoryless BECs with erasure probabilities δ_2 and δ_3, respectively. Using the notation of Sec. II-C, this means that Y_2 and Y_3 are related to X_1 via,

$$\begin{aligned} Y_2 &= X_1 + E_2 \\ Y_3 &= X_1 + E_3 \end{aligned} \tag{12}$$

where E_2 and E_3 are independent erasure noise variables, distributed as Erasure(δ_2) and Erasure(δ_3), respectively.

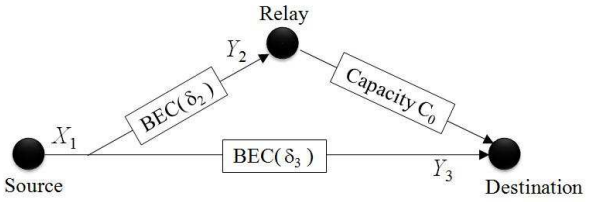


Fig. 4. The binary erasure relay channel.

Following [12], we assume that the relay is full-duplex, meaning that it can listen and transmit simultaneously. We define communication strategies and achievable rates in the standard way, see [12]. Specifically, the signal transmitted by the relay at time i may depend only on the channel outputs it observed at times $j = 1,...,i-1$.

Two relay communication strategies that frequently appear in the literature are decode-and-forward (DF), and compress-and-forward (CF) (see e.g. [32], [26], [12] as well as the tutorial [31]). With DF, the relay decodes the codeword from the source, and then cooperates with that node in the delivery of the associated message. CF focuses on scenarios where decoding of the codeword transmitted from the source is not

possible at the relay. An overview of this strategy was provided in Sec. I-C.

The achievable rates with DF and CF can be obtained using expressions in [27][15]. With DF, any rate R given by,

$$R \leq \min\Big(I(X_1;Y_2),\ I(X_1;Y_3)+C_o\Big)$$

is achievable, where the distribution $P_{X_1}(x_1)$ of X_1 is a parameter that can be optimized, and the distributions of the rest of the variables are derived from the channel transitions. The maximum achievable rate R_{DF} can be shown to equal,

$$R_{\text{DF}} \quad = \quad \min\Big(1-\delta_2, 1-\delta_3+C_o\Big) \tag{13}$$

Turning to CF, by [27] the following rates are achievable,

$$R \leq I(X_1;\ \hat{Y}_2, Y_3) \tag{14}$$

$\hat{Y}_2$ is an auxiliary random variable which is dependent on the relay output Y_2. The distributions $P_{X_1}(x_1)$ of X_1 and $P_{\hat{Y}_2|Y_2}(\hat{y}_2|y_2)$ of $\hat{Y}_2$ (conditioned on Y_2) are parameters which can be optimized. Their selection is constrained by the following condition, which needs to be satisfied:

$$I(Y_2;\hat{Y}_2 \mid Y_3) \leq C_o \tag{15}$$

Evaluation of the optimal choices for $P_{X_1}(x_1)$ and $P_{\hat{Y}_2|Y_2}(\hat{y}_2|y_2)$ is beyond the scope of our work. In this paper, we confine ourselves to X_1 which is uniformly distributed in $\{0,1\}$ and $\hat{Y}_2$ which is distributed as

$$\hat{Y}_2 = Y_2 + \hat{E}_2 \tag{16}$$

where $\hat{E}_2 \sim \text{Erasure}(\hat{\delta}_2)$ and is independent of Y_2. $\hat{E}_2$ corresponds to distortion, as applied by CF at the relay (see Sec. I-C).

Our choice for the distribution of $\hat{E}_2$ was guided by ease of analysis[16]. We will make similar choices later, in our design of methods based on soft-DF (Sec. III-C below), and so the comparison will be fair. With these choices, (14) and (15) become,

$$R \leq 1-(\delta_2 \circ \hat{\delta}_2)\cdot\delta_3 \tag{17}$$

$$h(\delta_2 \circ \hat{\delta}_2) + (1-\delta_2 \circ \hat{\delta}_2)\cdot\delta_3 - h(\hat{\delta}_2)\cdot(1-\delta_2) \leq C_o \tag{18}$$

where the operation $\circ$ is defined by (5). In the context of our discussion of distortion in Sec. I-C, $\hat{\delta}_2$ is the level of distortion in the signal conveyed by the relay to the destination. We define R_{CF} to equal R, as given by the right hand side of (17), evaluated at the minimal $\hat{\delta}_2$ which satisfies (18) (note that $\hat{\delta}_2 = 1$ renders the left-hand-side zero, and so the minimal $\hat{\delta}_2$ is well-defined). Explicitly,

$$R_{\text{CF}} = \max_{\hat{\delta}_2 \text{ satisfies (18)}} \Big\{1-(\delta_2 \circ \hat{\delta}_2)\cdot\delta_3\Big\} \tag{19}$$

The strategies described in [12], to achieve R_{DF} and R_{CF}, involve randomly generated codes (according to a uniform distribution in $\{0,1\}$), which are "good" for the point-to-point BEC channel[17]. In Sec. V-A we will use them as benchmarks, and provide examples of applications of soft-DF-BP that rely on "bad" codes, and outperform both these rates. In Sec. II-D we will further discuss related bounds on the performance of "good" codes.

B. Definition of Soft-DF-BP

As noted in Sec. I-C, soft-DF-BP is based on CF. With CF, the relay forwards its channel observation un-decoded to the destination. With soft-DF-BP, it first attempts to *estimate* the transmitted codeword, and forwards the resulting estimate to the destination. The destination combines this estimate with its own channel observation, and attempts to decode the source's codeword using both.

As also noted in Sec. I-C, the relay communicates its estimate to the destination using its channel to that node. To fit the capacity of this channel, the estimate is first compressed, to reduce the required bandwidth. Often, compression is not enough, and the signal needs to be distorted (lossy compression) to further reduce its entropy. Some reduction in the necessary rate can also be achieved without distortion, using a variant of Wyner-Ziv coding [64]. With this approach, the destination exploits the signal it obtained via its channel from the source, as side-information when reconstructing the relay's estimate (from the signal communicated by that node). Specifically, it relies on the statistical dependencies between the two signals.

A detailed discussion of compression at the relay and reconstruction at the destination, is provided by Cover and El Gamal [12, Theorem 6] in the context of communication using CF (the strategy's name having been coined later [32]). In this paper, we apply their results to communication using soft-DF-BP (see Sec. III-C and Appendix II below).

The details of soft-DF-BP are as follows.

Algorithm 2 (Soft-DF-BP):

- **Source.** Select a codeword $\mathbf{x}_1$ from an LDPC code $\mathcal{C}$ (which will be specified later), and transmit it over the channel.
- **Relay.**
  1) *Soft decoding.* Apply BP (Algorithm 1) to compute an estimate of $\mathbf{x}_1$ from the channel output $\mathbf{y}_2$ (see Sec. I-C for a discussion of the application of BP to estimation). The estimate is denoted $\mathbf{y}_2^{\text{BP}}$.
  2) *Wyner-Ziv compression.* Apply a vector quantizer to map $\mathbf{y}_2^{\text{BP}}$ to a distorted version $\hat{\mathbf{y}}_2^{\text{BP}}$, and communicate it to the destination (this will be elaborated in Sec. III-C).
- **Destination.**
  1) *Reconstruction.* Reconstruct $\hat{\mathbf{y}}_2^{\text{BP}}$ from the signal transmitted by the relay, using the output $\mathbf{y}_3$ of the channel from the source as side-information.
  2) *Decoding.* Apply BP to the vector $\hat{\mathbf{y}}_2^{\text{BP}} \cdot \mathbf{y}_3$ (i.e., use this vector instead of $\mathbf{y}$ in Algorithm 1), where

[15]The analysis of [27] specializes the results of [12] to settings where the outputs at the destination are decoupled, as in our formulation.

[16]Similar motivation guided the choice of auxiliary variables in [32, Sec. VII.A], in the context of CF over the Gaussian relay channel.

[17]More precisely, a sequence of codes generated in this way is "good" with probability 1, the probability implied by their random generation.

multiplication is defined as in Sec. II-C. The output of the algorithm is denoted $\mathbf{y}_3^{\text{BP}}$.

Remark 2: Recall from our definition of BP (Sec. II-E) that $\mathbf{y}_2^{\text{BP}}$, the output of BP at the relay, is defined over the alphabet $\{0,1,e\}$.
Ultimately, our measure of performance is the erasure rate (see Definition 2) at the output $\mathbf{y}_3^{\text{BP}}$ of soft-DF-BP at the destination.

C. Analysis Framework

Our analysis of soft-DF-BP will rely on a simple assumption, which involves the statistical relation between the relay estimate $\mathbf{Y}_2^{\text{BP}}$ and its distorted version $\hat{\mathbf{Y}}_2^{\text{BP}}$ (the random variables corresponding to $\mathbf{y}_2^{\text{BP}}$ and $\hat{\mathbf{y}}_2^{\text{BP}}$ as defined above). We begin by describing our assumption, and follow by a theorem which justifies its use. Formally, the justification will apply to a slightly modified version of soft-DF-BP (in comparison to Algorithm 2). We conjecture that in practice the modification is not needed, and our analysis results apply to the original version as well.

We model the statistical relation between the components of $\mathbf{Y}_2^{\text{BP}}$ and $\hat{\mathbf{Y}}_2^{\text{BP}}$ in the following way, which parallels (16) of our discussion of CF.

$$\hat{Y}_{2,i}^{\text{BP}} = Y_{2,i}^{\text{BP}} + \hat{E}_{2,i} \tag{20}$$

where $\hat{E}_{2,i}$ are i.i.d erasure noise components, $\hat{E}_{2,i} \sim \text{Erasure}(\hat{\delta}_2)$. We refer to the vector $\hat{\mathbf{E}}_2$ as the *quantization noise* vector. $\hat{\delta}_2$ is the quantization noise *level*, and will be discussed shortly. It is convenient to view (20) as a stochastic channel between $Y_{2,i}^{\text{BP}}$ and $\hat{Y}_{2,i}^{\text{BP}}$. We formulate this in the following definition.

Definition 5: In the *stochastic channel setup* for analysis of soft-DF-BP, Wyner-Ziv compression at the relay, and reconstruction at the destination, are replaced by transmission over a virtual BEC with erasure probability $\hat{\delta}_2$. We define the $(\delta_2, \delta_3, \hat{\delta}_2)$ *stochastic relay channel* as the channel that includes the physical links from the source to the relay and the destination (which are BECs with erasure probabilities δ_2 and δ_3, respectively), and the above virtual BEC.

We will occasionally refer to the $(\delta_2, \delta_3, C_o)$ channel of Sec. III-A as the *physical* relay channel, to distinguish it from the above stochastic one.

In our analysis, we will be interested in the erasure rate (Definition 2) at the output of the destination's BP decoder, in the above stochastic channel setup.

The following theorem provides the formal justification for our analysis.

Theorem 1: Let $(\delta_2, \delta_3, C_o)$ be the parameters of an erasure relay channel as defined in Sec. III-A. Let $\{\mathcal{C}_n\}_{n=1}^{\infty}$ be a sequence of LDPC codes with rate R, where n is the block length of $\mathcal{C}_n$, and let $\hat{\delta}_2, \epsilon \in [0,1]$. Assume the following conditions hold.

1) The erasure rate at the output of soft-DF-BP when applied in the $(\delta_2, \delta_3, \hat{\delta}_2)$ stochastic channel setup, is upper bounded by ϵ, with a probability that approaches 1 with n.
2) The following inequality is satisfied for large enough n,

$$\frac{1}{n} \cdot I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3) \leq C_o \tag{21}$$

Then a rate of $R \cdot (1 - h(\epsilon/R))$ is achievable using a modified version of soft-DF-BP, over the physical $(\delta_2, \delta_3, C_o)$ channel.

The proof of this theorem relies on the analysis of CF [12, Theorem 6] and is provided in Appendix II. The modified version of soft-DF-BP introduces an extra *outer* code[18] which is concatenated with the LDPC code $\mathcal{C}$, and replaces BP decoding at the destination with joint-typicality decoding. It preserves the main features of Algorithm 2, specifically soft decoding by BP at the relay, and Wyner-Ziv compression. Typically, we will be interested in negligibly small values of ϵ (e.g. 10^{-6}), and so the term $1 - h(\epsilon/R)$ will be very close to 1.

Recall that in our description of soft-DF-BP (Algorithm 2) we left out the details of Wyner-Ziv compression at the relay and reconstruction at the destination. In the modified version of the algorithm (Theorem 1), we assume the implementations described in the proof of [12, Theorem 6].

Equation (21) in the second condition in Theorem 1 parallels (15) from the analysis of CF. The distributions of the various random variables on the left hand side of (21) are obtained from the following discussion: $\hat{\mathbf{Y}}_2^{\text{BP}}$ is related to $\mathbf{Y}_2^{\text{BP}}$ via (20). $\mathbf{Y}_2^{\text{BP}}$ is a deterministic function of the random channel output at the relay $\mathbf{Y}_2$, being the output of an application of BP to this vector. $\mathbf{Y}_2$ and $\mathbf{Y}_3$ are both obtained from the transmitted codeword $\mathbf{X}_1$ via the channels from the source to the relay and destination, respectively. Finally, $\mathbf{X}_1$ is uniformly distributed within the LDPC code $\mathcal{C}_n$.

By our above discussion, given a $(\delta_2, \delta_3, C_o)$ erasure relay channel, an application of soft-DF-BP to the channel involves specifying not only the edge distributions (λ, ρ) for a sequence of LDPC codes (as in point-to-point communication) but also the quantization noise level $\hat{\delta}_2$. In our analysis, we will typically begin with a pair (λ, ρ), and select the quantization noise level by minimizing $\hat{\delta}_2$ subject to (21).

Our analysis now focuses on evaluating the two conditions of Theorem 1. The first condition will be discussed in Sections III-D and III-E and the second in Sec. III-F.

D. Background: Density Evolution

Our analysis of the bit erasure rate at the output of soft-DF-BP will rely on an extension of density evolution [47]. In this section we review density evolution as applied to the analysis of LDPC codes over point-to-point BECs, and examine the difficulty in extending it to soft-DF-BP over erasure relay channels. Overcoming this difficulty will be our focus in Sec. III-E. A complete and rigorous discussion of density evolution is available e.g. [36],[47],[48], and in this section we restrict our discussion to its essentials.

Density evolution is a numerical algorithm for approximating the bit erasure rate (Definition 2) at the output of BP (Algorithm 1, Sec. II-E). Its approximation is asymptotically

[18] A similar technique was applied in various contexts, e.g. [17, Theorem 3] implicitly relies on a similar derivation.

precise, in the sense that the realized erasure rate can be proven to approach it in probability, exponentially with the the LDPC block length n. Specifically, a concentration theorem relates the realized erasure rate to the probability that an *individual* message is an erasure. Density evolution computes an asymptotically-precise estimate of this probability.

Consider a rightbound message $\mathrm{r}_{ij}^{(\ell)}$ at iteration ℓ of BP (see Algorithm 1). This message is a function of leftbound messages that were computed in the preceding iteration, which in turn are functions of other messages. It is useful to depict this recursive structure in a *computation graph*, as shown in Fig. 5. The variable node which produced $\mathrm{r}_{ij}^{(\ell)}$ is drawn at the bottom of the graph. The check nodes $j' \in \mathcal{N}(i) \backslash \{j\}$ whose leftbound messages $\mathrm{l}_{j'i}^{(\ell)}$ were used in the computation of $\mathrm{r}_{ij}^{(\ell)}$ (see (9)) are drawn directly above node i. For each such check node j', the variable nodes whose rightbound messages were used in the computation of $\mathrm{l}_{j'i}^{(\ell)}$ are drawn above j', and so forth.

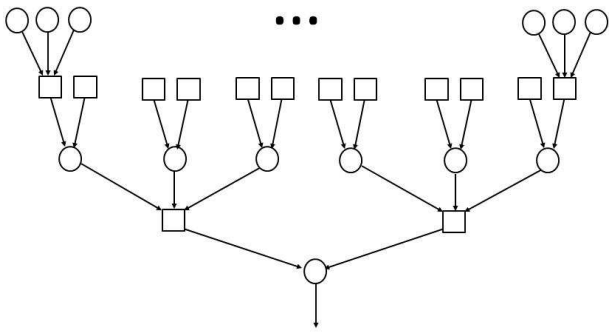


Fig. 5. The computation graph of a message $\mathrm{r}_{ij}^{(\ell)}$.

We now make some simplifying assumptions. First, in this section only, we confine our attention to regular LDPC codes (see Sec. II-E). In the following sections, we will allow irregular codes as well. The extension to such codes is possible using the concepts of [37],[47] and will not be elaborated. Second, we condition our analysis on the event that the all-zero codeword was transmitted by the source. Our interest in $\mathrm{r}_{ij}^{(\ell)}$ is confined to the question of whether or not it equals an erasure. An examination of BP (Algorithm 1) reveals that this only depends on the channel transitions, and not on the transmitted codeword, and thus our analysis will apply to other cases as well. Having fixed the transmitted codeword, the messages of BP, as well as its final output, are now functions of the realizations of the BEC channel transitions alone (or equivalently, the erasure noise components (4)). Lastly, we assume that the computation graph (Fig. 5) contains no loops. Assuming the code's Tanner graph was generated by the random method that was mentioned in Sec. II-E, this can be shown to hold, with high probability, at all but an exponentially small (in n) fraction of the messages at iteration ℓ [47].

With these assumptions, a number of observations can be made. First, computation graphs corresponding to different rightbound messages at iteration ℓ have identical structures. Since the channel transitions probabilities are also identical, this implies that the distribution of $\mathrm{R}_{ij}^{(\ell)}$ (the random variable corresponding to $\mathrm{r}_{ij}^{(\ell)}$) is independent of the edge (i,j). We denote this distribution by $P_R^{(\ell)}$. A similar argument can be made of the leftbound messages $\{\mathrm{l}_{j'i}^{(\ell)}\}$ (see (10)) on which the value of $\mathrm{R}_{ij}^{(\ell)}$ relies, and we denote their distributions by $P_L^{(\ell)}$.

Each leftbound message $\mathrm{l}_{j'i}^{(\ell)}$ is a function of the channel transitions corresponding to the variable nodes in the sub-graph spanning upward (in Fig. 5) from the check node that produced it. As the computation graph contains no loops, sub-graphs corresponding to different messages do not intersect, and their nodes are distinct. Since the channel transitions are independent (by the memorylessness of the channel), this implies that the variables $\{\mathrm{L}_{j'i}^{(\ell)}\}$ (the random variables corresponding to $\{\mathrm{l}_{j'i}^{(\ell)}\}$) are also statistically independent. Furthermore, they are independent of the channel output Y_i, whose value also affects $\mathrm{R}_{ij}^{(\ell)}$ in (4). These observations lead to simple equations [49, Sec. III.A], which can be used to compute ("evolve") $P_R^{(\ell)}$ from $P_L^{(\ell)}$ and P_Y (the distribution of Y_i). Similar arguments can be applied to compute $P_L^{(\ell)}$ from $P_R^{(\ell-1)}$ and so forth. These recursive equations are the basis for density evolution.

Unfortunately, these equations do not apply straightforwardly to the analysis of BP, as used by soft-DF-BP at the destination (they do apply to its analysis at the relay). The input to BP at the destination is $\hat{\mathbf{y}}_2^{\mathrm{BP}} \cdot \mathbf{y}_3$ (see Algorithm 2). Unlike the output of a memoryless BEC, the erasures in $\hat{\mathbf{Y}}_2^{\mathrm{BP}}$ are in general *not* statistically independent. In the context of our above discussion, this means that the leftbound messages $\{\mathrm{L}_{j'i}^{(\ell)}\}$ are now functions of *dependent* random variables, and are thus no longer independent.

E. Simultaneous Density Evolution

To overcome this difficulty (as noted in Sec. I-C), in this section we define a new algorithm which we call *simultaneous*-BP (sim-BP). Sim-BP plays a role equivalent to both applications of BP (at the relay and the destination) used by soft-DF-BP. Recall from Sec. III-C that the setup of our analysis assumes a stochastic relay channel (Definition 5). The input to sim-BP is a triplet of vectors, $(\mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{e}}_2)$, where $\mathbf{y}_2$ and $\mathbf{y}_3$ correspond to the relay and destination channel observations, and $\hat{\mathbf{e}}_2$ is the realization of the quantization noise vector. The output of sim-BP is a pair of estimates of the codeword $\mathbf{x}_1$ that was transmitted by the source.

As noted in Sec. I-C, sim-BP cannot be realized, because the relay and destination are physically separated, and so combined access to both their channel observations ($\mathbf{y}_2$ and $\mathbf{y}_3$) is not possible. The assumption that the algorithm has access to the quantization noise $\hat{\mathbf{e}}_2$ is similarly unusual. Sim-BP is thus intended only as a theoretical tool for analysis. We will prove that its output is degraded (in the sense of Definition 3) with respect to the of output of soft-DF-BP, and thus its performance can be used to bound that of soft-DF-BP. Most importantly, its structure will enable rigorous analysis using a variation of density evolution, which we will call *simultaneous density evolution* (sim-DE).

Recall from Sec. III-D that our difficulty in applying density evolution involved the vector $\hat{\mathbf{y}}_2^{\text{BP}} \cdot \mathbf{y}_3$, which at the destination replaces $\mathbf{y}$ of the definition of BP (Algorithm 1). The components of the vector appear in the expression for the rightbound message (9). We rewrite this expression below, explicitly as it is used at the destination.

$$\mathrm{r}_{ij}^{(3,\ell)} = \begin{cases} \hat{y}_{2,i}^{\text{BP}} \cdot y_{3,i}, & \ell = 0, \\ \hat{y}_{2,i}^{\text{BP}} \cdot y_{3,i} \cdot \prod_{j' \in \mathcal{N}(i) \setminus \{j\}} \mathrm{l}_{j'i}^{(3,\ell)}, & \ell > 0. \end{cases} \tag{22}$$

where the superscript $(3,\ell)$ denotes messages at the destination (node 3) at BP iteration ℓ. By (20), the components of $\hat{\mathbf{y}}_2^{\text{BP}}$ are modeled by,

$$\hat{y}_{2,i}^{\text{BP}} = y_{2,i}^{\text{BP}} + \hat{e}_{2,i} \tag{23}$$

To overcome the difficulties involving $\hat{\mathbf{y}}_2^{\text{BP}} \cdot \mathbf{y}_3$, with sim-BP we replace (22) with,

$$\mathrm{r}_{ij}^{(3,\ell)} = \begin{cases} \hat{\mathrm{r}}_{ij}^{(2,\ell)} \cdot y_{3,i}, & \ell = 0, \\ \hat{\mathrm{r}}_{ij}^{(2,\ell)} \cdot y_{3,i} \cdot \prod_{j' \in \mathcal{N}(i) \setminus \{j\}} \mathrm{l}_{j'i}^{(3,\ell)}, & \ell > 0. \end{cases} \tag{24}$$

where,

$$\hat{\mathrm{r}}_{ij}^{(2,\ell)} = \mathrm{r}_{ij}^{(2,\ell)} + \hat{e}_{2,i} \tag{25}$$

where $\mathrm{r}_{ij}^{(2,\ell)}$ is the rightbound message computed by the *relay's* BP decoder and $\hat{e}_{2,i}$ is the same realization of the erasure noise as in (23) (which, as mentioned above, is included in the input to the algorithm).

More precisely, sim-BP operates as follows. It is a message-passing algorithm, and alternates between leftbound and rightbound iterations. Its messages are *pairs*, denoted $(\mathrm{r}_{ij}^{(2,\ell)}, \mathrm{r}_{ij}^{(3,\ell)})$ and $(\mathrm{l}_{ji}^{(2,\ell)}, \mathrm{l}_{ji}^{(3,\ell)})$. Components $\mathrm{r}_{ij}^{(2,\ell)}$ and $\mathrm{l}_{ji}^{(2,\ell)}$ are identical to the messages exchanged with soft-DF-BP by the relay's BP decoder. That is, they are functions of the relay channel observation $\mathbf{y}_2$. The expressions by which they are computed are obtained from (9) and (10) (replacing y_i, $\mathrm{r}_{ij}^{(\ell)}$ and $\mathrm{l}_{ji}^{(\ell)}$ with $y_{2,i}$, $\mathrm{r}_{ij}^{(2,\ell)}$ and $\mathrm{l}_{ji}^{(2,\ell)}$). Components $\mathrm{r}_{ij}^{(3,\ell)}$ and $\mathrm{l}_{ji}^{(3,\ell)}$ parallel messages exchanged with soft-DF-BP by the destination's BP decoder, but are not identical to them. $\mathrm{r}_{ij}^{(3,\ell)}$ is computed using (24). $\mathrm{l}_{ij}^{(3,\ell)}$ is computed using (10) (replacing $\mathrm{r}_{ij}^{(\ell)}$ and $\mathrm{l}_{ji}^{(\ell)}$ with $\mathrm{r}_{ij}^{(3,\ell)}$ and $\mathrm{l}_{ji}^{(3,\ell)}$). While this expression for $\mathrm{l}_{ij}^{(3,\ell)}$ is identical to the one used with soft-DF-BP, the dependence on $\mathrm{r}_{i'j}^{(3,\ell)}$ means that it too differs from the equivalent soft-DF-BP messages. We briefly summarize this description below.

Algorithm 3 (Simultaneous Belief Propagation (sim-BP)):

1) **Iterations.** Perform the following steps, alternately, a pre-determined t times.
   - *Rightbound iteration number* $\ell = 0, ..., t-1$. At all edges (i,j) compute a rightbound pair $(\mathrm{r}_{ij}^{(2,\ell)}, \mathrm{r}_{ij}^{(3,\ell)})$ in the following way: $\mathrm{r}_{ij}^{(2,\ell)}$ is computed by (9), making substitutions as mentioned above. $\mathrm{r}_{ij}^{(3,\ell)}$ is computed by (24).
   - *Leftbound iteration number* $\ell = 1, ..., t$. At all edges (j,i) compute a leftbound pair $(\mathrm{l}_{ji}^{(2,\ell)}, \mathrm{l}_{ji}^{(3,\ell)})$ in the following way: Both components are computed by (10), making substitutions as mentioned above.
2) **Final decisions.** For each $i = 1, ..., n$ compute a pair $(y_{2,i}^{\text{BP}}, y_{3,i}^{\text{BP}})$ as follows: $y_{2,i}^{\text{BP}}$ is computed as in BP (Algorithm 1), expression (11), replacing y_i and $\mathrm{l}_{ji}^{(t)}$ with $y_{2,i}$ and $\mathrm{l}_{ji}^{(2,t)}$. $y_{3,i}^{\text{BP}}$ is computed using the same expression, but replacing y_i and $\mathrm{l}_{ji}^{(t)}$ with $\hat{y}_{2,i}^{\text{BP}} \cdot y_{3,i}$ and $\mathrm{l}_{ji}^{(3,t)}$, where $\hat{y}_{2,i}^{\text{BP}}$ is computed by (23).

Note that while sim-BP outputs a pair of vectors $(\mathbf{y}_2^{\text{BP}}, \mathbf{y}_3^{\text{BP}})$, in practice we are only interested in $\mathbf{y}_3^{\text{BP}}$, which is an estimate of the transmitted source codeword. The following theorem relates it to the output of the destination's BP decoder with soft-DF-BP.

Theorem 2: Consider an instance of communication using soft-DF-BP in a stochastic channel setup (Definition 5). Let $\mathbf{y}_2, \mathbf{y}_3$ denote the channel observations, and $\hat{\mathbf{e}}_2$ denote the quantization noise by which $\hat{\mathbf{y}}_2^{\text{BP}}$ and $\mathbf{y}_2^{\text{BP}}$ are related as (20). Let $\mathbf{y}_3^{\text{BP}}$ denote the output of BP at the destination. Let $\mathbf{y}'^{\text{BP}}_3$ denote the output of sim-BP when provided with precisely the same vectors $(\mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{e}}_2)$. Then $\mathbf{y}'^{\text{BP}}_3$ is degraded with respect to $\mathbf{y}_3^{\text{BP}}$ in the sense of Definition 3.

By this theorem, we can use sim-BP to upper bound the erasure rate at the output of soft-DF-BP at the destination. The theorem makes intuitive sense, because the messages $\{\mathrm{r}_{ij}^{(2,\ell)}\}$, which sim-BP uses in (24), are intermediate values, computed in the process of BP, while components $y_{2,i}^{\text{BP}}$ which the destination's BP decoder uses in (22), are final decisions, whose quality is expected to be better. This argument will be made rigorous in Appendix III-A.

Sim-BP preserves the essential features of BP which make its analysis using density evolution possible. Namely, it is a message-passing algorithm, and its inputs $(\mathbf{y}_2, \mathbf{y}_3, \hat{\mathbf{e}}_2)$ are vectors of independently distributed components (conditioned on the transmission of the all-zero codeword). Simultaneous density evolution (sim-DE) tracks the quantities $P_R^{(\ell)}(x_2, x_3)$ and $P_L^{(\ell)}(x_2, x_3)$, corresponding to the joint probability functions of message pairs $(\mathrm{R}_{ij}^{(2,\ell)}, \mathrm{R}_{ij}^{(3,\ell)})$ and $(\mathrm{L}_{ji}^{(2,\ell)}, \mathrm{L}_{ji}^{(3,\ell)})$ (uppercase denotes random variables), respectively.

The inputs to sim-DE are a triplet $(\delta_2, \delta_3, \hat{\delta}_2)$ and a pair (λ, ρ), that characterize the stochastic relay channel (Definition 5) and the LDPC code (Sec. II-E), respectively. The details of the algorithm are provided in Appendix III-B. The algorithm concludes by outputting $P^{(\text{Final})}(x_2, x_3)$, corresponding to the distribution of a final decision pair $(Y_{2,i}^{\text{BP}}, Y_{3,i}^{\text{BP}})$. We let $P_e^{(\text{Final})}$ denote the probability that $Y_{3,i}^{\text{BP}} = e$, as evaluated by computing the appropriate marginal distribution from $P^{(\text{Final})}(x_2, x_3)$. The following theorem relates this value to the performance of sim-BP.

Theorem 3: Consider communication in the stochastic channel setup (Definition 5). Let $(\mathbf{Y}_2, \mathbf{Y}_3, \hat{\mathbf{E}}_2)$ be the random relay and destination channel observations and the quantization noise. Assume the code $\mathcal{C}$ used was selected at random from a (λ, ρ) LDPC ensemble of block length n (see Sec. II-E). Let $(\mathbf{Y}_2^{\text{BP}}, \mathbf{Y}_3^{\text{BP}})$ denote the output of sim-BP, when provided with $(\mathbf{Y}_2, \mathbf{Y}_3, \hat{\mathbf{E}}_2)$ as inputs, and applied to the Tanner graph of $\mathcal{C}$. Then for any $\epsilon > 0$ and large enough n, the following holds,

$$\Pr\left[\left| P_e(\mathbf{Y}_3^{\text{BP}}) - P_e^{(\text{Final})} \right| > \epsilon\right] < \mathrm{e}^{-\beta\epsilon^2 n}$$

where, $P_e^{(\text{Final})}$ is as defined above, $P_e(\mathbf{Y}_3^{\text{BP}})$ is as in Definition 2, and $\beta > 0$ is some constant, which is independent of n and ϵ.

The theorem implies that the erasure rate at the output of sim-BP approaches sim-DE's prediction in probability, exponentially in n. The proof follows in direct lines as the proof of [47, Theorem 2] and is omitted.

Combined with Theorem 2, the theorem gives us an upper bound on the erasure rate at the output of soft-DF-BP. In the context of the analysis strategy of Sec. III-C, this addresses the first of the two conditions of Theorem 1.

F. Analysis of the Quantization Noise $\hat{\mathbf{E}}_2$

We now examine the second condition of Theorem 1. In our analysis in Sec. V-A, this condition will determine the level $\hat{\delta}_2$ of the quantization noise, which we will choose as the minimal value still satisfying (21).

In the evaluation of $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$ (the left hand side of (21)), the distributions of the variables involved (see Sec. III-C) are implicitly functions of the code $\mathcal{C}$ in use, through their dependence on the transmitted $\mathbf{X}_1$, which is randomly distributed in $\mathcal{C}$. When evaluating $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$, the code $\mathcal{C}$ is assumed to be fixed. In our analysis, however, $\mathcal{C}$ is randomly selected from an ensemble (see Sec. II-E). The value of $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$ is thus a random variable whose value is determined by $\mathcal{C}$.

We begin with a naive bound on $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$. This bound closely resembles the left hand side of (18), which was the evaluation of $I(Y_2; \hat{Y}_2 | Y_3)$ in the context of communication using CF. The proof of the lemma is provided in Appendix IV.

Lemma 1 (Naive bound): Let $\mathcal{C}$ be selected at random from a (λ, ρ) LDPC ensemble with block length n. Then the following holds for large enough n, with probability at least $1 - \exp(-\alpha n^{1/3})$ (the probability being over the random selection of $\mathcal{C}$),

$$\frac{1}{n} I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3) \leq h(\delta_2^{\text{BP}} \circ \hat{\delta}_2) + \\ + (1 - \delta_2^{\text{BP}} \circ \hat{\delta}_2) \cdot \delta_3 - h(\hat{\delta}_2) \cdot (1 - \delta_2^{\text{BP}}) + o(1) \quad (26)$$

where δ_2^{BP} is the expected erasure rate at the output of the relay's BP decoder (Sec. III-B) as computed by density evolution (Sec. III-D). $o(1)$ is some function of n, dependent on λ, ρ and t (the number of relay BP iterations) that approaches zero with n. $\alpha > 0$ is a constant similarly dependent on λ, ρ and t.

The above bound, however, does not exploit strong dependencies that often exist between the components of the vector $\mathbf{Y}_2^{\text{BP}}$ (as mentioned in Sec. III-D), which result from the simple structures of LDPC codes. The following theorem exploits these dependencies to produce a stronger bound. Unlike Lemma 1, our bound in this theorem applies to right-regular LDPC codes only (see Sec. II-E).

Theorem 4: Let $\mathcal{C}$ be selected at random from a right-regular LDPC ensemble (λ, d) with block length n. Then the following holds for large enough n, with probability at least $1 - \exp(-\alpha n^{1/3})$,

$$\frac{1}{n} I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3) \quad \leq \quad I^+(\hat{\delta}_2) + o(1) \quad (27)$$

where α and $o(1)$ are defined as in Lemma 1,

$$I^+(\hat{\delta}_2) = \text{l.d.f.} \left[\min \left(I_1^+(\hat{\delta}_2),\ I_2^+(\hat{\delta}_2) \right) \right] \quad (28)$$

where l.d.f.$[f]$ denotes the largest descending function that is upper bounded by $f(\cdot)$. That is,

$$\text{l.d.f.}[f](x) = \inf_{t \leq x} f(t)$$

and,

$$I_1^+(\hat{\delta}_2) \quad = \quad A(\hat{\delta}_2) + h(\delta_2^{\text{BP}} \circ \hat{\delta}_2) \quad (29)$$

$$I_2^+(\hat{\delta}_2) \quad = \quad A(\hat{\delta}_2) + f(\delta_2^{\text{BP}}) + (1 - \delta_2^{\text{BP}}) \cdot h(\hat{\delta}_2) \quad (30)$$

where δ_2^{BP} is defined as in Lemma 1, $A(\hat{\delta}_2)$ and $f(\alpha)$ are provided by equations (31) and (32) on the following page.

The proof of the theorem is provided in Appendix V. In the proof, we make use of the following dependencies between the components of $\mathbf{Y}_2^{\text{BP}}$.

1) **Dependencies between bits discovered by BP:** Each component $Y_{2,i}^{\text{BP}}$ that corresponds to a bit that was erased by the channel but discovered at some BP iteration, is dependent and equal to the sum (modulo-2) of $d-1$ components $Y_{2,i_1}^{\text{BP}}, ..., Y_{2,i_{d-1}}^{\text{BP}}$ other bits. This is best seen by examining the following simplified algorithm, due to Luby $et\ al.$ [36, Algorithm 1], which is equivalent to BP. Like BP, this algorithm is iterative, and relies on the Tanner graph representation of the LDPC code. However, it is not a message-passing algorithm.

Algorithm 4 (Simplified-BP):

1) *Initialization.* Set the value of each variable node to the channel output.
2) *Iteration* $\ell = 1, ..., t$. Perform the following operation simultaneously at all check nodes: At check node j, if the values at *all but one* of the adjacent variable are known (not erased), set the remaining unknown variable node to the modulo-2 sum of the others.

In [48, Problem 3.10] this algorithm was shown to yield precisely the same output as BP. Therefore, we may assume without loss of generality that it is the one applied by the relay of soft-DF-BP. Each bit i uncovered by this algorithm (and by extension, BP) is clearly equal to the sum of $d-1$ bits $i_1, ..., i_{d-1}$ that were revealed in previous iterations, or by the channel. Equivalently stated, the entropy of $Y_{2,i}^{\text{BP}}$, given $Y_{2,i_1}^{\text{BP}}, ..., Y_{2,i_{d-1}}^{\text{BP}}$ is zero. A good lossy compression scheme for $\mathbf{Y}_2^{\text{BP}}$ can exploit this and spend less rate to describe $Y_{2,i}^{\text{BP}}$, under the assumption that with high probability, the decoder will have access to all $Y_{2,i_1}^{\text{BP}}, ..., Y_{2,i_{d-1}}^{\text{BP}}$ (this is made rigorous in Appendix V).

2) **Dependencies between erasures at the output of BP:** To adequately communicate $\mathbf{Y}_2^{\text{BP}}$ to the destination (or its distorted version, $\hat{\mathbf{Y}}_2^{\text{BP}}$, see Sec. III-B), the relay must be able to approximately convey the locations (time indices) of $\mathbf{Y}_2^{\text{BP}}$ components that are equal to erasure. Unlike the output of a memoryless BEC, these locations are not arbitrary. Di $et\ al.$ [14, Lemma 1.1] proved that they correspond to a *stopping set* of the code $\mathcal{C}$ (see [14] for its definition). Typically, the number of stopping sets of a given size is significantly smaller than the number of similar-sized (arbitrary)

$$A(\hat{\delta}_2) = \delta_3(1-\hat{\delta}_2)\left[(1-\delta_2)+(\delta_2-\delta_2^{\text{BP}})\left(1-(1-\hat{\delta}_2)^{d-1}\right)\right]-(1-\delta_2^{\text{BP}})\cdot h(\hat{\delta}_2) \quad (31)$$

$$f(\alpha) = \max_{\beta\in(0,\gamma)}\left[\log\inf_{x>0,y>0}\left(\frac{\prod_i(1+xy^i)^{\tilde{\lambda}_i}}{x^\alpha y^\beta}\right)+\log\inf_{x>0}\left(\frac{\left[(1+x)^d-d\cdot x\right]^{(1-R)}}{x^\beta}\right)-\gamma\cdot h\left(\frac{\beta}{\gamma}\right)\right] \quad (32)$$

$$\gamma \triangleq \sum_i i\cdot\tilde{\lambda}_i$$

subsets of $\{1,...,n\}$. Thus, for the relay to describe to the destination the locations of the erasures in $\mathbf{Y}_2^{\text{BP}}$, it can conserve rate by providing the serial number of a stopping set (given some enumeration of the stopping sets) rather than describing the precise indices. This is exploited by the bound implied by (30) above.

In Fig. 9 (Sec. V-A) we have plotted the bounds of Lemma 1 and Theorem 4 for a specific LDPC code ensemble, as well as a bound that applies to "good" codes. A discussion of the bounds is provided in that section.

G. Limitations of "Good" Codes

In Sec. III-A, we mentioned R_{DF} and R_{CF} as benchmarks for the performance of "good" codes. Unlike our analysis in Sec. IV-C (in the context of interference channels), we were *not* able to obtain tight bounds that apply in general to *any* relay strategy that involves "good" codes. In this section, we nonetheless examine soft-DF-BP, and point out limitations on its components, when the code $\mathcal{C}$ it relies on (see Algorithm 2) is taken from a sequence of "good" codes. We also provide bounds on the achievable rates in this setting, which hold under certain plausible assumptions. We conjecture that in practice, soft-DF-BP, as well as other relay strategies, are restricted by similar bounds, when confined to "good" codes.

In our analysis below, we distinguish between two ranges for the code rate R: $R \leq 1-\delta_2$ and $R > 1-\delta_2$. In the first range, R is below the capacity of the source-relay link. As we will see below, in this range "good" codes enjoy an advantage over "bad" ones, in terms of soft decoding at the relay. Specifically, complete decoding is possible, and thus methods like DF (see Sec. III-A), which involve decoding at the relay, are also possible. However, the rates in this range are limited. We define the upper bound,

$$R_{\text{DF}-\text{UB}} = 1-\delta_2 \quad (33)$$

to be the maximum rate in this range.

Our main focus is on the second range, $R > 1-\delta_2$. In this range, R exceeds the capacity of the source-relay link, and thus complete decoding at the relay is not possible. Methods like CF, however, which do not involve complete decoding, are potentially possible. In our development of soft-DF-BP, in Sec. III-B, our objective was to improve upon CF. We will argue that this is unlikely to be possible when soft-DF-BP is used with "good" codes.

We begin by considering soft decoding, as applied by soft-DF-BP at the relay. Our analysis will rely on the following theorem, which examines the estimation error at a destination node of a *point-to-point* BEC. The theorem focuses on $P_{\text{MAP}}(\mathcal{C};\,\delta)$, the expected bit erasure rate (Definition 2) at the output of a maximum *a posteriori* (MAP) decoder for a linear code[19] $\mathcal{C}$, when used over a BEC(δ).

Theorem 5: Let $\{\mathcal{C}_n\}_{n=1}^{\infty}$ be a sequence of linear codes, of rate R, which is "good" for the BEC (see Definition 4). Let $\delta^\star = 1-R$ be the BEC Shannon limit for rate R (Definition 1). Then the following holds,

$$\lim_{n\to\infty} P_{\text{MAP}}(\mathcal{C}_n;\,\delta) = \begin{cases} \delta, & \delta > \delta^\star; \\ 0, & \delta < \delta^\star. \end{cases} \qquad \forall \delta\in[0,1], \delta\neq\delta^\star \quad (34)$$

The results of this theorem resemble the ones that were presented in Fig. 1. Specifically, at high values of δ (paralleling *low* SNRs in Fig. 1), MAP decoding of "good" codes collapses, and its output closely resembles the raw channel signal at its input. The proof of the theorem is a variation of the proof of [45][Equation (14)] and is provided in Appendix VI-A. It relies on the relationship between mutual information and input estimates, which was recently discovered in several contexts (see Palomar and Verdú [44], Méasson $et\ al.$ [39] and Ashikhmin $et\ al.$ [2]).

The setting of Theorem 5 is more general than required for an analysis of soft-DF-BP. It applies to arbitrary "good" linear codes[20], rather than "good" LDPC codes, and to MAP decoding (estimation) rather than BP decoding. The application to soft BP decoding of "good" LDPC codes at a relay, is obtained as a corollary, in the following way. We will apply the theorem to examine the source-relay link. As noted above, our focus is on rates in the range $R > 1-\delta_2$, or equivalently on erasure probabilities $\delta_2 > 1-R = \delta^\star$ (where $\delta^\star$ is as defined in Theorem 5). This is precisely the range where MAP estimation of "good" codes collapses. By the optimality of MAP decoding, the expected erasure rate at the output of BP cannot be lower than $P_{\text{MAP}}(\mathcal{C};\delta_2)$, and thus BP estimation collapses too.

It is still possible that BP estimation achieves a reduction in the number of erasures which is *sub*-linear in the block length, and that this reduction produces a meaningful benefit. We conjecture that this is not the case. A more detailed analysis is deferred to later work. Formally, we make the following conjecture.

[19]MAP decoding of linear codes over the BEC produces vectors over the alphabet $\{0,1,e\}$ (see e.g. [48, Sec. 3.2.1]). More precisely, the bitwise *a-posteriori* probability $P_{X_i|\mathbf{Y}}(1|\mathbf{y})$ on which it relies can be shown to belong to the set $\{0,1,1/2\}$, indicating complete confidence (0 or 1) in the decoded bit or complete lack of it (1/2). In some formulations of the algorithm, a random decision is made when $P_{X_i|\mathbf{Y}}(1\mid\mathbf{y})=1/2$. In this paper, we assume 1/2 is mapped to e, producing the desired alphabet.

[20]Our choice to focus on linear codes was made for simplicity. We believe that the result can easily be extended to arbitrary "good" codes.

Conjecture 1: Consider the asymptotic performance of soft-DF-BP, assuming that the code $\mathcal{C}$ in Algorithm 2 was taken from a sequence of "good" codes. Then an analysis that assumes that the output $\mathbf{Y}_2^{\text{BP}}$ of BP at the relay is identical to the channel output $\mathbf{Y}_2$, involves no loss in generality. Similarly, we may assume $\hat{\mathbf{Y}}_2^{\text{BP}} = \hat{\mathbf{Y}}_2$, where the components of $\hat{\mathbf{Y}}_2$ and $\hat{\mathbf{Y}}_2^{\text{BP}}$ are given by (16) and (20), respectively.

We now proceed to consider Wyner-Ziv compression as applied by soft-DF-BP at the relay. By our discussion in Sec. III-C (following Theorem 1), in our analysis of soft-DF-BP we select the quantization noise level by minimizing $\hat{\delta}_2$ subject to (21). In Sec. III-A, the left hand side of (15), which is the CF equivalent of (21), was evaluated to equal the left hand side of (18) (subject to (16)). In Theorem 4 (Sec. III-F) we have seen that with "bad" codes we can do better, and the left hand side of (21) is upper bounded by a value that is often lower than the left hand side of (18). We now wish to show that "good" codes do *not* exhibit a similar advantage. This is partially addressed by Theorem 6 below. Note that in this theorem we focus on $1/n \cdot I(\mathbf{Y}_2; \hat{\mathbf{Y}}_2 \,|\, \mathbf{Y}_3)$ rather than $1/n \cdot I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} | \mathbf{Y}_3)$ as in (21), following Conjecture 1 above.

Theorem 6: Let $\hat{\delta}_2 > 0$, and let $\{\mathcal{C}_n\}_{n=1}^{\infty}$ and R be defined as in Theorem 5. Assume,

$$R = 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3 \tag{35}$$

Then the following holds:

$$\frac{1}{n} \cdot I(\mathbf{Y}_2; \hat{\mathbf{Y}}_2 \,|\, \mathbf{Y}_3) = h(\delta_2 \circ \hat{\delta}_2) + \\ +(1 - \delta_2 \circ \hat{\delta}_2) \cdot \delta_3 - h(\hat{\delta}_2) \cdot (1 - \delta_2) + o(1) \tag{36}$$

where $o(1)$ approaches zero with n (n being the block length of $\mathcal{C}_n$, see Remark 1).

The proof of the theorem is provided in Appendix VI-B.

We would now like to determine the rates that are achievable with soft-DF-BP, when confined to "good" codes. While Theorem 1 (Sec. III-C) provides a set of achievable rates, the conditions of the theorem have only been proven to be sufficient, but not necessary for a rate to be achievable. We nonetheless make the following conjecture.

Conjecture 2: The conditions of Theorem 1 are necessary as well as sufficient for a rate to be achievable with soft-DF-BP.

The conditions of Theorem 1 are closely related to the conditions of CF (14) and (15) (Sec. III-A). Specifically, the first condition of the theorem involves performance over the $(\delta_2, \delta_3, \hat{\delta}_2)$ stochastic relay channel (Definition 5), whose outputs are $\hat{\mathbf{Y}}_2$ and $\mathbf{Y}_3$ (where we have substituted $\hat{\mathbf{Y}}_2^{\text{BP}}$ by $\hat{\mathbf{Y}}_2$ as explained above). The achievable rates in this setting are bounded by the capacity of the channel[21] from $\mathbf{X}_1$ to $(\hat{\mathbf{Y}}_2, \mathbf{Y}_3)$, which is $1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3$. The first condition of Theorem 1 thus implies (17) (which was evaluated from (14), see Sec. III-A). Following Conjecture 1, we replace the second condition with the following inequality,

$$\frac{1}{n} \cdot I(\mathbf{Y}_2; \hat{\mathbf{Y}}_2 \,|\, \mathbf{Y}_3) \leq C_o \tag{37}$$

By the above discussion, the rates achievable with soft-DF-BP, when confined to "good" sequences of codes, subject to Conjectures 1 and 2, are upper-bounded by,

$$R_{\text{UB}} = \sup_{\text{"good" } \{\mathcal{C}_n\}} \left\{ \begin{array}{c} R = R_{\{\mathcal{C}_n\}} : \exists \hat{\delta}_2 \in [0,1] : \\ R \leq 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3, \\ \text{(37) holds for large} \\ \text{enough } n. \end{array} \right\} \tag{38}$$

The supremum is over all "good" code sequences, and $R_{\{\mathcal{C}_n\}}$ denotes the rate of the sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$.

Finally, in Appendix VI-C we apply Theorem 6 to show that R_{UB} is upper bounded by R_{CF}. In Sec. V-A we will rely on $R_{\text{DF}-\text{UB}}$ and R_{CF} as benchmarks for the rates that are achievable with soft-DF-BP, when confined to "good" code sequences.

IV. Coding for the Symmetric BIAWGN Interference Channel

A. Channel Model and Achievable Strategies

Fig. 6 depicts the (h, σ) symmetric BIAWGN interference channel. The channel transition probabilities are defined by the following equations,

$$Y_1 = X_1 + h \cdot X_2 + Z_1 \\ Y_2 = h \cdot X_1 + X_2 + Z_2 \tag{39}$$

where Y_1 and Y_2 are the channel outputs at the two destinations (respectively), X_1 and X_2 and are the transmitted signals. Unlike typical formulations of AWGN interference channels (e.g. [18]) we restrict X_1 and X_2 to $\{\pm 1\}$. Z_1 and Z_2 are statistically independent zero-mean real-valued Gaussian random variables with variance σ^2, whose realizations at different time instances are also independent. h and σ are positive constants, known to all nodes, and h is restricted to the range $h \in (0, 1)$ (i.e. we are interested in weak interference scenarios [11]).

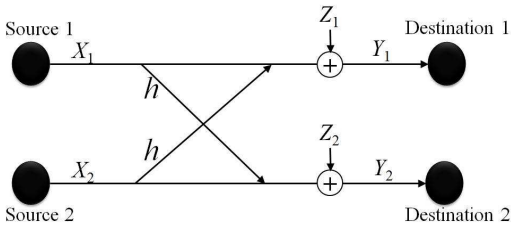


Fig. 6. The symmetric BIAWGN interference channel.

We define achievable strategies for this channel in the standard way (e.g. [18, Sec. II]). Theoretic analysis of interference channels typically focuses on the capacity region, i.e. the set of pairs (R_1, R_2) such that communication at rate R_1 (resp. R_2) is achievable between the first (resp. second) source-destination pair. In this paper we confine our attention to

[21] Note that in this discussion, we focus on performance in terms of frame errors (rather than bit errors). In the context of Theorem 1, this is equivalent to setting $\epsilon = 0$.

achievable *symmetric* rates. That is, values $R > 0$ such that the pair (R, R) is achievable. In our discussion, the following terminology will be useful.

Definition 6: When considering one of the two destinations of an interference channel, we refer to the corresponding source, which produced the message that the destination must decode, as the *primary* source. The other source is called the *interfering* source.

Two sub-optimal communication strategies that frequently appear as benchmarks in literature on interference channels, are multi-user detection (MUD) and single-user detection (SUD). With MUD, each destination attempts to decode the interference, as well as the desired primary signal (i.e. produced by the primary source). With SUD, each destination treats the interference as noise, alongside the channel noise. Achievable rates with both strategies can be evaluated from the following expressions.

$$R_{\text{MUD}} = \min\left(I(X_2;Y_1 \mid X_1),\ \frac{1}{2}I(X_1,X_2;\ Y_1)\right) \quad (40)$$

$$R_{\text{SUD}} = I(X_1;Y_1) \quad (41)$$

While in general, the distributions $P_{X_1}(x_1)$ of X_1, and $P_{X_2}(x_2)$ of X_2, are parameters to be optimized, in this paper we confine our attention to X_1 and X_2 are uniformly distribution in $\{\pm 1\}$. With these choices, (40) and (41) can be evaluated numerically. Note that X_1 and X_2 are assumed to be independently distributed. Additionally, on the right hand side of (41), the distribution of Y_1 is implicitly a function of $P_{X_2}(x_2)$ (as well as $P_{X_1}(x_1)$) by virtue of (39). The expressions (40) and (41), are easily obtained from analysis of multiple-access channels [13, Theorem 14.3.3], and point-to-point (single-user) channel capacity [13, Theorem 8.7.1], respectively.

The strategies described in [13], to achieve rates (40) and (41), involve randomly generated codes (according to a uniform distribution in $\{\pm 1\}$), which are "good" for the point-to-point BIAWGN channel[22]. Furthermore, in Sec. IV-C we will show that they are in fact the *best* that can be done in communication with *any* "good" code sequence. In Sec. V-B we will thus use them as benchmarks for communication using "good" codes.

B. Definition and Analysis Framework for Soft-IC-BP

We assume the two sources use LDPC codes, whose identities will be discussed later. Soft-IC-BP is applied at each of the two destinations of the interference channel.

At each destination, soft-IC-BP attempts to decode the primary codeword (see Definition 6), and produces an estimate of the interference as a byproduct. As noted in Sec. I-C, the algorithm coincides with iterative-MUD as defined e.g. by [6], [1], [51] (and references therein). Unlike standard applications of iterative-MUD, however, in applications of soft-IC-BP we tolerate a large error in the estimation of the interference (but not the primary codeword).

[22]More precisely (as in Sec. III-A), a sequence of codes generated in this way is "good" with probability 1, the probability implied by their random generation.

We now provide a general overview of soft-IC-BP and of its analysis methods. A complete discussion is available in the above-mentioned references. Our description is intended to point out specific features of our implementation, and also as background for the optimization of LDPC codes, which will be discussed in Sec. V-B and Appendix VIII-B.

It is convenient to perceive the operation of soft-IC-BP at each destination, as the parallel operation of two decoders, the first decoding the primary codeword, and the second estimating the interference. The two decoders iteratively exchange information to improve their respective performance. At each iteration, the information each decoder obtains from the other, assists it in better canceling the signal produced by the other source, in order to better estimate its own signal.

More precisely, soft-IC-BP progresses through the exchange of messages between the nodes of a *factor graph* [33], which represents the communication setting (see Fig. 7) at the destination. This graph contains the Tanner graphs of the primary and interference LDPC codes (see Sec. II-E), as well as additional nodes, including n *state* nodes[23]. Each state node corresponds the received signal at one time instance. It is linked to one variable node from each Tanner graph, each corresponding to a transmitted bit from one of the two sources at the given time instance. Decoding includes standard LDPC decoding iterations (see e.g. [47]) as well as variable-to-state and state-to-variable iterations, which implement an exchange of information between the two Tanner graphs. For precise details regarding the computation of the messages see [51][Sec. 2].

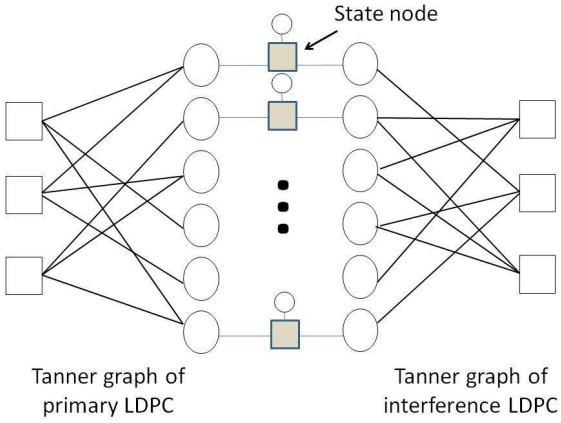


Fig. 7. An example of the factor graph for an application of soft-IC.

In this paper, we have adopted a number of attributes of the design of [51]. Namely, we have assumed that the LDPC codes used by the two sources have the same block lengths and edge distributions (λ, ρ). Under this assumption, the number of nodes of any given degree within the Tanner graphs of the codes is the same. We further assumed that the nodes are arranged so that the two variable nodes that are linked to each state node have the same degree. In [51, Sec. 4] this is known as the *no-interleaver hypothesis*. Lastly, we assumed *parallel*

[23]In [51] they are called "state-check" nodes.

scheduling. This means that decoding iterations at both Tanner graphs are computed in parallel.

Analysis of soft-IC-BP is possible by an application of density evolution, similar to the one that was discussed in Sec. III-D, in the context of BP over the BEC. Once again, density evolution tracks the distributions of messages exchanged at the various iterations of soft-IC-BP. In addition to rightbound and leftbound LDPC iterations, the algorithm tracks the distributions of variable-to-state and state-to-variable messages. A distinction is made between the messages exchanged in the Tanner graph of the primary codeword, and the messages in the graph of the interference, whose distributions are expected to be different. Unlike BP over the BEC, the messages of soft-IC-BP are taken from a large, continuous alphabet (the real number field), and so their distributions as tracked by density evolution are defined over this alphabet (more precisely, a fine grid over the real-number field as [47]).

A detailed discussion of density evolution for iterative-MUD is available in [6, Sec. IV.A] and [48, Sec. 5.5]. Like [48], our computation of the evolution of distributions through variable-to-state and state-to-variable iterations is performed precisely, rather than by Monte-Carlo simulations as in [6]. Unlike [6, Sec. IV.A] and [48, Example 5.34], our reliance on the above-mentioned no-interleaver hypothesis implies that the degrees of the variable node linked to each state node are *not* independent (in fact they are equal). In our implementation of density evolution, we account for this by considering the variable-to-state, state-to-variable and following rightbound iteration as a combined single iteration.

As in the case of standard density evolution, a concentration theorem exists [6, Proposition 1] that asserts that the realized bit error rate, with both the primary and interference codewords, approaches density evolutions's prediction in probability, exponentially in the block length n.

C. Limitations of "Good" Codes

We now turn our attention to limitations on the performance of "good" codes. In Sec. IV-A, we mentioned that the MUD and SUD achievable strategies both rely on communication with "good" codes. The following theorem shows that their achievable rates ((40) and (41)) bound the performance of *any* communication strategy that relies on *any* set of "good" codes. These results are stronger than the ones we obtained in Sec. III-G, in the context of erasure relay channels.

Theorem 7: Consider communication over a symmetric BI-AWGN interference channel. Assume the two sources use equal block length codes taken from "good" code sequences $\{\mathcal{C}_{1,n}\}_{n=1}^{\infty}$ and $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$, respectively, which have rate R (see Sec. II-D). Assume the probabilities of decoding error, under maximum-likelihood decoding, at both destinations, approach zero with the block length n. Then the following holds,

$$R \leq \max\left(R_{\text{MUD}}, R_{\text{SUD}}\right) \tag{42}$$

The proof of this theorem is provided in Appendix VII. The proof builds on the converse of the capacity theorem of multiple-access channels, see e.g. [13, Sec. 14.3.4]. For example, consider the setting facing Destination 1. Once the destination has decoded the primary codeword (see Definition 6), it is able to subtract it. The remaining signal is equivalent to the output of a point-to-point BIAWGN channel, whose input is the interference $\mathbf{X}_2$. If R is lower than the capacity of this channel, then by the "goodness" of $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$, complete decoding of the interference is possible. Thus, the communication setting in this case resembles a multiple-access scenario, leading to the bound $R \leq R_{\text{MUD}}$ (see Appendix VII for the rigorous details).

If R is greater than the capacity of the above point-to-point BIAWGN channel, then complete decoding of the interference $\mathbf{X}_2$ is not possible. However, relying on the "goodness" of $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$, we are still able to bound its entropy given the channel output $\mathbf{Y}_1$ and the primary codeword. We apply this bound in Appendix VII to show that in this case, $R \leq R_{\text{SUD}}$ must hold.

Remark 3: Note that in communication using a code $\mathcal{C}$, we mean that the source simply maps each message to a codeword of $\mathcal{C}$, and does not manipulate $\mathcal{C}$, e.g. by combining it with another code, as in the Han-Kobayashi strategy [25].

Finally, in Sec. V-B below, we provide examples of simple-structured "bad" LDPC codes, which are capable of communication at rates that exceed R_{MUD} and R_{SUD}, thus surpassing the performance of "good" codes.

V. Code Design Methods and Numerical Results

A. Erasure Relay

As noted in Sec. III-C, application of soft-DF-BP to a $(\delta_2, \delta_3, C_o)$ erasure relay channel involves specifying the edge distributions (λ, ρ) for the LDPC code, as well as the quantization noise level $\hat{\delta}_2$, such that the conditions of Theorem 1 are satisfied (for ϵ which will be specified later). Given parameters $(\lambda, \rho, \hat{\delta}_2)$ of such an application, we use sim-DE, as defined in Sec. III-E to verify that the first condition of that theorem is satisfied, and the bounds of Theorem 4 to verify the second condition. Our objective is to maximize the communication rate, as measured by the design rate (8) corresponding to (λ, ρ). As benchmarks for comparison, we use R_{DF}, R_{CF} and $R_{\text{DF}-\text{UB}}$, defined by (13), (19) and (33), following our discussions in Sections III-A and III-G.

To design effective soft-DF-BP parameters, we applied a semi-heuristic hill climbing algorithm based on Richardson *et al.* [49, sec. IV.A] and [48, Example 4.139]. Once the parameters were obtained, they were verified by the *non*-heuristic methods described above. Our algorithm starts with an initial *admissible* $(\lambda, \rho, \hat{\delta}_2)$ triplet, i.e. one that satisfies the conditions of Theorem 1. It proceeds by iteratively attempting to improve it, so that at each iteration the design rate is increased, and the triplet is still admissible. The details of the algorithm are provided in Appendix VIII-A.

We designed codes for an erasure relay channel with parameters $\delta_2 = 0.5$, $\delta_3 = 0.82$ and $C_o = 0.9$ (C_o is measured in bits, see Sec. II-A). We applied the above design procedure and obtained edge distributions (λ, ρ) corresponding to a design

rate of $R = 0.5056$. The parameters of the codes are,

$$\lambda_{2,3,4,23,24,100} = (0.2289, 0.04532, 0.2361, 0.233, 0.03178, 0.2249), \quad \rho_{10} = 1, \quad \hat{\delta}_2 = 0.212 \tag{43}$$

As the degrees of the check-nodes in such (λ, ρ) codes are bounded (and equal to 10), code sequences corresponding to (λ, ρ) are "bad" for the point-to-point BEC channel[24]. By Theorem 4, the right-hand-side of (21) is upper bounded by 0.9 for large enough n, and so the second condition of Theorem 1 is satisfied. The erasure rate at the output of soft-DF-BP, as predicted by sim-DE, is upper bounded by $\epsilon = 1.54 \cdot 10^{-5}$ in probability with the block length n. By Theorem 1, this means that the achievable rate with such codes is $R = 0.5053$.

Our benchmarks (see above) for this channel are $R_{\text{DF}} = 0.5$, $R_{\text{CF}} = 0.49867$ and $R_{\text{DF}-\text{UB}} = 0.5$. All benchmarks were surpassed by the above rate R which is achievable by soft-DF-BP. By our discussion in Sec. III-G, R_{CF} and $R_{\text{DF}-\text{UB}}$ are likely to upper bound the achievable rates with any implementation of soft-DF-BP that applies "good" codes. Thus, our gap from the benchmarks indicates the advantage of using "bad" codes[25].

It is interesting to examine additional aspects of the performance of the "bad" LDPC code sequences corresponding the above (λ, ρ) pair, in comparison with the performance of "good" code sequences. In Fig. 8, we consider communication over a point-to-point BEC. This figure parallels Fig. 1 (Sec. I). The various curves correspond to the erasure rates at the output of estimation algorithms at the destination, as functions of the the channel's erasure probability. The first curve corresponds to the asymptotic erasure rate at the output of MAP estimation with "good" codes of rate 0.5053. The value plotted is the asymptotic limit as determined in Theorem 5. The second curve corresponds to the the expected erasure rate at the output of BP estimation, when applied to communication using a randomly generated (λ, ρ) LDPC code (where (λ, ρ) are defined by (43)), as predicted by density evolution (see Sec. III-D). The last curve corresponds to uncoded communications, where estimation cannot improve upon the raw channel output.

The erasure probability of the source-relay link of the above-mentioned $(\delta_2, \delta_3, C_o)$ channel is 0.5. At this point, the "good" codes curve evaluates to 0.5, while the (λ, ρ) LDPC curve achieves an erasure rate of 0.3016. Thus, with (λ, ρ) LDPC, soft-DF-BP forwards a much better signal $\mathbf{y}_2^{\text{BP}}$ (see Alg. 2) to the destination, than when "good" codes are applied.

Another factor that affects the achievable rates at the destination, is the level of quantization noise $\hat{\delta}_2$. In Fig. 9 we have plotted the required $\hat{\delta}_2$ as a function of the capacity C_o of the relay-destination link. In all curves, δ_2 and δ_3 are fixed and equal to the values specified above. The first curve in Fig. 9 corresponds to communication using "good" codes, and was computed by minimizing $\hat{\delta}_2$ subject to (18). $\hat{\delta}_2$ equals the level of noise in applications of CF (see Sec. III-A). By the discussion in Sec. III-G, it is also likely to be a lower bound on the noise level in applications of soft-DF-BP that involve "good" codes[26]. The second curve in Fig. 9 corresponds to communication using the a randomly selected code from the (λ, ρ) LDPC ensemble (where (λ, ρ) are defined by (43)), and was computed by minimizing $\hat{\delta}_2$ subject to $I^+(\hat{\delta}_2) \leq C_o$, where $I^+(\hat{\delta}_2)$ is given by (28). The third curve corresponds to the naive upper bound of Lemma 1, and is similarly computed using (26) (ignoring the $o(1)$ term).

The capacity of the relay-destination link in the above-mentioned $(\delta_2, \delta_3, C_o)$ channel is 0.9. At this point, the "good" codes curve evaluates to $\hat{\delta}_2 = 0.223$, while our LDPC codes require at most $\hat{\delta}_2 = 0.212$. Thus, communication of the signal $\mathbf{y}_2^{\text{BP}}$ from the relay to the destination, is possible with a lower level of distortion when our above "bad" LDPC codes are used, than when "good" codes are applied. It is also interesting to observe the performance of the curve at higher values of C_o. With our above LDPC code, lossless ($\hat{\delta}_2 = 0$) compression is possible when $C_o = 0.9463$. The "good" codes curve requires $C_o = 1.41$.

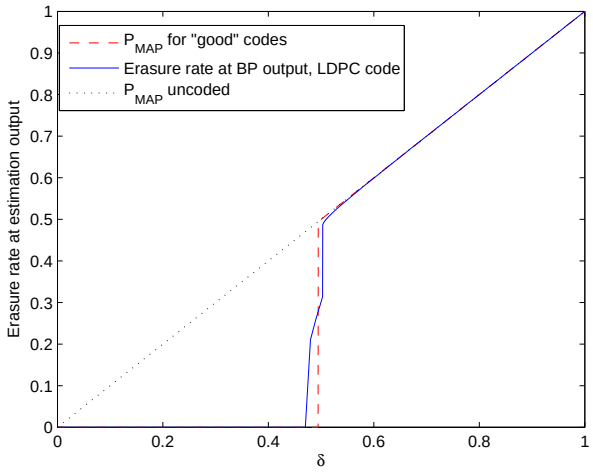


Fig. 8. Erasure rates as functions of the BEC erasure probability δ.

B. Symmetric BIAWGN Interference

An application of soft-IC-BP for a (h, σ) symmetric BI-AWGN interference channel involves specifying the edge distributions (λ, ρ) for the LDPC codes in use. As noted in Sec. IV-B, we assume both sources use the same edge distributions. We use density evolution, as discussed in Sec. IV-B, to verify the codes' performance. Our objective is again to maximize the design rate (8), corresponding to (λ, ρ). As benchmarks for comparison, we use R_{MUD} and R_{SUD}, as defined by (40) and (41) in Sec. IV-A.

[24]This follows from the discussion of Sec. I-B, by observing that in the parity-check matrix that corresponds to the Tanner graph of such (λ, ρ) LDPC codes (see e.g. [8, Sec. II.A]), the average weight of each row is 10.

[25]We also experimented with partial-DF (see Sec. I-A). Unfortunately, we were not able to design an application of the strategy whose performance exceeds the above rates achieved by DF and CF. A similar difficulty was reported by [31, Sec. 4.2.7] in the context of full-duplex AWGN channels. Further optimization of partial-DF is beyond the scope of this work.

[26]Strictly speaking, the level of quantization noise may be lower, because Theorem 6 requires $R = 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3$ while communication may be possible if $R \leq 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3$. However, by the discussion in Appendix VI-C, the achievable rates remain unaltered even if we assume the curve indeed lower bounds the required $\hat{\delta}_2$.

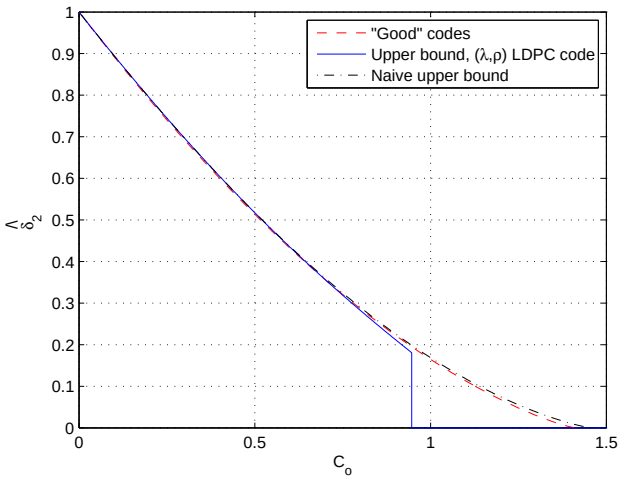


Fig. 9. Quantization noise $\hat{\delta}_2$ as a function of the capacity of relay-destination link C_o.

As in our above discussion of soft-DF-BP, we apply a variation of the semi-heuristic hill-climbing algorithm of [49] to design effective codes. In adapting the algorithm, special care was taken to avoid attraction to local maxima that are particular to our setting. The details of the algorithm are provided in Appendix VIII-B.

We designed codes for a symmetric BIAWGN interference channels with parameters $h = 0.839$ and $\sigma = 1.075$. We applied the above design procedure and obtained edge distributions (λ, ρ) below.

$$\lambda_{2,3,10,11,55,56,57} = (0.2949, 0.2036, 0.05943, 0.0001219, 0.2399, 0.09542, 0.1065), \quad \rho_6 = 1 \tag{44}$$

Once again, the degrees of the check-nodes in such (λ, ρ) codes are bounded (and equal to 6), and so code sequences corresponding to (λ, ρ) are "bad" for the point-to-point BI-AWGN channel (this follows by the same arguments as in Sec. V-A above). The bit error rate at the output of soft-IC-BP, as predicted by density evolution, approaches $4 \cdot 10^{-6}$ in probability with the block length n. This figure refers to decoding at each destination, of the codeword transmitted by the corresponding source. The bit error rate in decoding of the codeword sent from the interfering source, approaches 0.062 with n. By the symmetry of the problem, these figures are identical at both destinations.

Our benchmarks for the above channel are $R_{\text{MUD}} = 0.3237$ and $R_{\text{SUD}} = 0.308$ (measured in bits, see Sec. II-A). By Theorem 7 these rates upper bound the achievable rates of applications of "good" codes. The design rate for the code specified by (44) is 0.3243. This rate exceeds both our benchmarks, indicating the potential of "bad" codes. The above-mentioned bit error rate of 0.062, in the decoding of the interfering codeword, is greater than zero but lower than the bit error rate with bitwise decoding (i.e., when the code structure is not exploited), which equals 0.301. Thus, partial decoding of the interference was achieved.

As noted in Sec. I-A, the best-known rates for the interference channel are achieved by the Han-Kobayashi (HK) strategy [25], which like soft-IC-BP, involves partial decoding. In Appendix VIII-C we describe an application of HK for the above channel which is provably capable of communication at rate 0.333. This rate exceeds our above application of soft-IC-BP. As noted in Sec. I-B, however, this comes at a price with respect to practical considerations like decoding complexity. We also verify in Appendix VIII-C that the codes used by HK to achieve this performance are point-to-point "bad". This reinforces our insight that "bad" codes have an inherent role in methods that rely on partial decoding.

VI. Conclusion

Multi-terminal communications poses a much richer research problem than traditional (point-to-point) communications. While coding for point-to-point channels needs only consider the performance at the destination, multi-terminal channels offer additional degrees of freedom, by enabling partial decoding at non-destination nodes (e.g. relays) as well.

The approach of partial-DF [12] and HK [25] (see Sec. I), which involves manipulating randomly-generated codes, is a natural extension of the traditional analysis of point-to-point channels. Over such channels, randomly-generated codes were the first "good" codes to be found. Soft-DF and soft-IC, by comparison, often rely on simple-structured codes, which over point-to-point channels have been shown to be suboptimal ("bad"). In this paper, we have demonstrated that such codes may in fact offer benefits in multi-terminal scenarios, which are intrinsically related to their point-to-point "badness". Our main contribution has been a rigorous analysis of these benefits, in terms of achievable communication rates.

Many open problems remain. Most important, in our view, are the questions posed in Sec. I-B. Specifically, an analysis of the tradeoff between achievable rates and computational complexity is of great practical interest. Tightening of our bounds (e.g. Theorem 4) and refinement of our LDPC optimization algorithm (Appendix VIII-A) may enable the improvement of the achievable rates reported in Sec. V, and yield insight on the capacities of relay and interference channels. Extensions of our results to additional relay and interference channels, as well as additional network models, are also interesting research problems. Specifically, we conjecture that the significance of partial decoding increases with the number of network nodes.

Our focus in this paper on LDPC codes was guided exclusively by ease of analysis. Our results, however, open the door to a re-evaluation of other "bad" (non capacity achieving) point-to-point codes, like simple convolutional, Reed-Muller and Reed-Solomon codes, in multi-terminal scenarios.

Appendix I
Details of the Curves in Figure 1

The MMSE values plotted in Fig. 1 correspond to the asymptotic normalized MMSE of a *sequence* codes. More precisely, given a code $\mathcal{C}_o$, we define,

$$\text{mmse}(\mathcal{C}_o; \text{SNR}) \triangleq \mathbb{E}\left[\|\hat{\mathbf{X}}(\mathbf{Y}) - \mathbf{X}\|^2 \right] \tag{45}$$

where $\mathbf{X}$ is the transmitted codeword, which is randomly distributed in $\mathcal{C}_o$, $\mathbf{Y}$ is the channel output at the destination and $\hat{\mathbf{X}}(\mathbf{Y})$ is the MMSE estimate of $\mathbf{X}$ given $\mathbf{Y}$. Note that we assume no distinction is made between information and parity bits, and estimation of the values of *all* transmitted code bits is performed.

The curves in Fig. 1 corresponds to the normalized MMSE of *sequences* of codes. Given a sequence $\mathcal{C} = \{\mathcal{C}_n\}_{n=1}^{\infty}$, we define its normalized MMSE as,

$$\text{mmse}(\mathcal{C}, \text{SNR}) \triangleq \lim_{n\to\infty} \frac{1}{n}\text{mmse}(\mathcal{C}_n, \text{SNR})$$

The curve in Fig. 1 corresponding to uncoded communications was evaluated as [23][Equation (17)]. In the curve corresponding to "good" code-sequences, we have defined "goodness" as Definition 4 with respect to BIAWGN channels. In the range SNR $<$ SNR* where SNR* is the Shannon limit for rate 1/2, we have relied on the analysis of [45][Equation (14)] to evaluate the curve. In the range SNR $>$ SNR*, we have relied on an analysis similar to the one in Appendix VI-A, with respect to estimation over the BEC, in the range $\delta < \delta^\star$.

The LDPC(2,4) curve corresponds to a sequence of codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$, where $\mathcal{C}_n$ was selected at random from the LDPC(2,4) ensemble of block length n (see Sec. II-E). The bound on the MMSE was explained in our paper [4, Sec. III.A].

APPENDIX II
PROOF OF THEOREM 1

We begin by an overview of our proof. Our main technique involves focusing on the *virtual* channel, obtained by encapsulating soft decoding, as performed at the relay with soft-DF-BP, into the channel model. This channel is identical to the relay channel of Sec. III-A with the exception that the relay output $Y_{2,i}$ at time i is replaced by $Y_{2,i}^{\text{BP}}$. Having encapsulated soft decoding into the channel model, the remaining components of soft-DF-BP now closely resemble CF. Thus, it would appear that we can apply the results of [12, Theorem 6] to analyze its performance.

However, the application of the above technique involves addressing two issues. First, unlike the relay channel output $\mathbf{Y}_2$, the erasures in $\mathbf{Y}_2^{\text{BP}}$ are in general not statistically independent. This means that the channel from components $X_{1,i}$ to $Y_{2,i}^{\text{BP}}$ is not memoryless, thus violating an assumption of [12, Theorem 6]. Second, the proof of [12, Theorem 6] only guarantees that $\hat{\mathbf{Y}}_2^{\text{BP}}$, as obtained at the destination, is *strongly typical* to (20). Our analysis of LDPC codes will require that its actual distribution match (20).

These issues are easily addressed by considering a modified version of soft-DF-BP that uses a code $\mathcal{C}^\star$, obtained by concatenating $\mathcal{C}$ (see Algorithm 2) with an outer code $\mathcal{C}_{\text{out}}$. Each communication thus involves multiple transmissions of codewords from the $\mathcal{C}$ (the *inner* code). $\mathcal{C}^\star$ replaces $\mathcal{C}$ at all operations of the modified algorithm (e.g. encoding at the source), with the exception of soft decoding at the relay, which is applied independently to each of the concatenated codewords of $\mathcal{C}$. We also assume that the destination applies joint-typicality decoding (see [12]) to decode $\mathcal{C}^\star$, rather than BP decoding.

Analysis is now simplified by examining the transition probabilities of the virtual *outer* channel. The source input alphabet of the channel consists of the codewords of $\mathcal{C}$, the relay output is $\mathbf{Y}_2^{\text{BP}}$, the relay-destination channel has capacity $n \cdot C_o$, and the output at the destination is $\mathbf{Y}_3$. Communication using modified soft-DF-BP involves multiple uses of this channel.

The virtual outer channel is clearly memoryless, thus addressing the first issue above. Furthermore, modified soft-DF-BP, as defined over this channel, now matches CF as defined in [12, Theorem 6]. Analysis of the joint-typicality decoder on which it relies is possible with the standard information-theoretic techniques used by [12], thus removing the second issue above. The analysis of [12, Theorem 6] (as specialized in [27], see Sec. III-A above) guarantees that if (21) holds, we can select $\mathcal{C}_{\text{out}}$ such that the rate

$$R_{\text{modified}} = \frac{1}{n} \cdot I(\mathbf{X}_1; \mathbf{Y}_3, \hat{\mathbf{Y}}_2^{\text{BP}}) \tag{46}$$

is achievable, where we assume that $\mathbf{X}_1$ is uniformly distributed in $\mathcal{C}$, and normalization by n is required because we are measuring rate in bits per use of the *inner* channel.

Finally, to evaluate (46), we apply the analysis of soft-DF-BP in the stochastic channel setup. That is, we consider a *formal* scenario, where the assumptions of the setup hold. Relying on Condition 1 of the theorem, the following inequality can now be shown to hold,

$$\frac{1}{n} \cdot I(\mathbf{X}_1; \mathbf{Y}_3, \hat{\mathbf{Y}}_2^{\text{BP}}) > R \cdot \Big(1 - h(\epsilon/R)\Big) + o(1) \tag{47}$$

where R is the rate $\mathcal{C}$ and $o(1)$ is a term that approaches zero with n. The proof of (47) relies on concepts similar to the proof of the joint source-channel coding theorem (see e.g. [38, Sec. 10.5]) and is omitted. The argument ϵ/R to the entropy function (rather than ϵ as in [38]) compensates for the fact that ϵ is the fraction of erroneous code bits rather than information bits as in [38].

□

APPENDIX III
RESULTS FOR SEC. III-E

A. Proof of Theorem 2

The proof relies on the properties of erasure multiplication and addition as defined (5) and (6). Specifically, it is easy to verify that if x' is degraded with respect to x and y' is degraded with respect to y, then $x'+y'$ is degraded with respect to $x+y$ and $x' \cdot y'$ is degraded with respect to $x \cdot y$. Similarly, x is degraded with respect to $x \cdot y$ for all $x, y \in \{0, 1, e\}$. We proceed with the following lemma

Lemma 2: Consider an instance of the application of BP (Algorithm 1) over the point-to-point BEC. Let $\text{r}_{ij}^{(\ell)}$ be a rightbound message computed at some intermediate iteration $\ell = 0, ..., t-1$, and y_i^{BP} the final decision later computed at the node i that produced $\text{r}_{ij}^{(\ell)}$. Then $\text{r}_{ij}^{(\ell)}$ is degraded with respect to y_i^{BP}.

Proof: We will actually prove a stronger result: $\mathrm{r}_{ij}^{(\ell-1)}$ is degraded with respect to $\mathrm{r}_{ij}^{(\ell)}$ for all $\ell = 1, ..., t-1$, and $\mathrm{r}_{ij}^{(t-1)}$ is degraded with respect to y_i^{BP}. The desired result will follow by the obvious transitivity of degradedness.

Our proof follows by induction on the iteration number ℓ. We start by comparing $\mathrm{r}_{ij}^{(0)}$ and $\mathrm{r}_{ij}^{(1)}$. By (9), $\mathrm{r}_{ij}^{(0)}$ equals the channel output y_i while $\mathrm{r}_{ij}^{(1)}$ is obtained by multiplying y_i with some other components. The result now follows by the above-mentioned property of erasure multiplication.

We proceed to examine $\mathrm{r}_{ij}^{(\ell-1)}$ and $\mathrm{r}_{ij}^{(\ell)}$ for $\ell = 2, ..., t-1$. By (9), both are functions of leftbound messages across the same edges, but at different iterations ($\{\mathrm{l}_{j'i}^{(\ell-1)}\}_{j' \in \mathcal{N}(i)\backslash\{j\}}$ and $\{\mathrm{l}_{j'i}^{(\ell)}\}_{j' \in \mathcal{N}(i)\backslash\{j\}}$, respectively), as well as the same y_i. If we could prove that each message $\mathrm{l}_{j'i}^{(\ell-1)}$ is degraded with respect to the corresponding $\mathrm{l}_{j'i}^{(\ell)}$, the result would follow by the above-mentioned property of erasure multiplication. Each such leftbound message is computed by (10). Again, both are functions of rightbound messages across the same edges, but at different iterations ($\{\mathrm{r}_{i'j'}^{(\ell-2)}\}_{i' \in \mathcal{N}(j')\backslash\{i'\}}$ and $\{\mathrm{r}_{i'j'}^{(\ell-1)}\}_{i' \in \mathcal{N}(j')\backslash\{i'\}}$, respectively). By induction, each message $\mathrm{r}_{i'j'}^{(\ell-2)}$ is degraded with respect to $\mathrm{r}_{i'j'}^{(\ell-1)}$ and the result now follows by the above-mentioned property of erasure addition.

Finally, the proof of the degradedness of $\mathrm{r}_{ij}^{(t-1)}$ with respect to y_i^{BP} follows by similar arguments and is omitted. □

We now introduce some notation. By construction, the first components $\mathrm{r}_{ij}^{(2,\ell)}$ and $\mathrm{l}_{ji}^{(2,\ell)}$ of the message pairs computed by sim-BP are identical to the messages computed by the relay's BP decoder with soft-DF-BP. The same does *not* hold for the other components of sim-BP and soft-DF-BP's messages at the destination. We let $\mathrm{r}_{ij}^{(3,\ell)}$ and $\mathrm{l}_{ij}^{(3,\ell)}$ denote the messages computed with soft-DF-BP at the destination, and $\mathrm{r}'^{(3,\ell)}_{ij}$ and $\mathrm{l}'^{(3,\ell)}_{ji}$ the corresponding components of sim-BP's message pairs.

The proof proceeds by showing that $\mathrm{r}'^{(3,\ell)}_{ij}$ is degraded with respect to $\mathrm{r}_{ij}^{(3,\ell)}$ for all ℓ, i, j. With the above notation, $\mathrm{r}'^{(3,\ell)}_{ij}$ is computed by (24), replacing $\mathrm{r}_{ij}^{(3,\ell)}$ and $\mathrm{l}_{ij}^{(3,\ell)}$ by $\mathrm{r}'^{(3,\ell)}_{ij}$ and $\mathrm{l}'^{(3,\ell)}_{ji}$, respectively. $\mathrm{r}_{ij}^{(3,\ell)}$ is computed by (22). By Lemma 2, $\mathrm{r}_{ij}^{(2,\ell)}$ is degraded with respect to $y_{2,i}^{\text{BP}}$ for all ℓ, i, j. By (23) and (25) and the above-mentioned properties of multiplication, it follows that $\hat{\mathrm{r}}_{ij}^{(2,\ell)}$ is degraded with respect to $\hat{y}_{2,i}^{\text{BP}}$. Each message $\mathrm{l}'^{(3,\ell)}_{j'i}$, $j' \in \mathcal{N}(i)\backslash\{j\}$ can be shown to be degraded with respect to $\mathrm{l}_{j'i}^{(3,\ell)}$ by induction, using similar arguments to the ones used in the proof of Lemma 2 above. The desired degradedness of $\mathrm{r}'^{(3,\ell)}_{ij}$ with respect to $\mathrm{r}_{ij}^{(3,\ell)}$ now follows by the above-mentioned properties of multiplication.

Finally, the proof of degradedness of $y'^{\text{BP}}_{3,i}$ with respect to $y_{3,i}^{\text{BP}}$ now follows from the above results by similar arguments and is omitted.

□

B. Details of Simultaneous Density Evolution

The description below relies on the discussion of Sec. III-E. The algorithm is based on the concepts of density evolution over point-to-point channels, as described in Sec. III-D. Like density evolution, it relies on the all-zero codeword assumption, and thus the distributions $P_R^{(\ell)}(x_2, x_3)$ and $P_L^{(\ell)}(x_2, x_3)$ that it tracks are confined to the range $\{0, e\} \times \{0, e\}$.

Algorithm 5 (Simultaneous Density Evolution (sim-DE)):

1) **Iterations.** Perform the following steps, alternately, a pre-determined t times.
   - *Rightbound iteration number* $l = 0, ..., t-1$. Set $P_R^{(\ell)} = \Gamma(\overline{P}_R^{(\ell)})$ where,
   $$\overline{P}_R^{(\ell)} = \begin{cases} P_{\delta_2} \cdot P_{\delta_3}, & \ell = 0, \\ \sum_i \lambda_i \cdot \left[\overline{P}_R^{(0)} \odot \left(P_L^{(\ell)} \right)^{\odot(i-1)} \right], & \ell > 0. \end{cases} \tag{48}$$
   where $P_\delta(\cdot)$ is defined for $\delta \in [0, 1], x \in \{0, e\}$ by,
   $$P_\delta(x) \triangleq \begin{cases} \delta, & x = e \\ 1-\delta, & x = 0 \end{cases}$$
   $P_{\delta_2} \cdot P_{\delta_3}$ is defined by (51) on the following page and the operation $\odot$ is defined by (52). $P^{\odot i} \triangleq P \odot P \odot ... \odot P$, i.e., the repeated application of the operation $\odot$ a number i times on P. Addition and multiplication by λ_i in (48) are performed componentwise (see (54)). Lastly, $\Gamma(\cdot)$ is defined by (55).
   - *Leftbound iteration number* $\ell = 1, ..., t$. $P_L^{(\ell)}$ is obtained by,
   $$P_L^{(\ell)} = \sum_j \rho_j \cdot \left[\left(P_R^{(\ell-1)} \right)^{\oplus(j-1)} \right] \tag{49}$$
   where the operation $\oplus$ is defined by (53), and where P_1 and P_2 are probability functions over $\{0, e\}^2$. $P^{\oplus i}$ is defined in the same way as $P^{\odot i}$
2) **Final Decisions.** Set $P^{(\text{Final})} = \Gamma(\overline{P}^{(\text{Final})})$ where,
   $$\overline{P}^{(\text{Final})} = \sum_i \tilde{\lambda}_i \cdot \left[\overline{P}_R^{(0)} \odot \left(P_L^{(\ell)} \right)^{\odot i} \right] \tag{50}$$
   where $\tilde{\lambda}_i$ is as defined in Sec. II-E.

Sim-DE follows the same concepts of density evolution as developed by Richardson *et al.* [47] and discussed in Sec. III-D. Its computations follow the expressions for sim-BP. Like standard density evolution, the incoming message pairs at each node, on which the computations for the outgoing pairs rely, are assumed to be mutually independent (although the components within each such pair are in general dependent). At variable nodes, the pairs are also assumed to be independent of the node's channel outputs $(Y_{2,i}, Y_{3,i}, \hat{E}_{2,i})$. These assumptions are justified by similar arguments to the ones in Sec. III-D, relying on the fact that conditioned on the transmission of the all-zero codeword, components $(Y_{2,i}, Y_{3,i}, \hat{E}_{2,i})$ corresponding to different indices i, are mutually independent.

The computations in a rightbound iteration have been simplified by introducing an intermediate step. Rather than determine $P_R^{(\ell)}$ directly, the algorithm first computes an auxiliary

$$P = P_{\delta_2} \cdot P_{\delta_3} \quad \Longleftrightarrow \quad P(x_2, x_3) = P_{\delta_2}(x_2) \cdot P_{\delta_3}(x_3) \qquad \forall x_2, x_3 \in \{0, e\} \tag{51}$$

$$P = P_1 \odot P_2 \quad \Longleftrightarrow \quad P(x_2, x_3) = \sum_{\substack{x_2^1, x_3^1, x_2^2, x_3^2 \in \{0,e\} \\ x_2^1 \cdot x_2^2 = x_2,\, x_3^1 \cdot x_3^2 = x_3}} P_1(x_2^1, x_3^1) \cdot P_2(x_2^2, x_3^2) \quad \forall x_2, x_3 \in \{0, e\} \tag{52}$$

$$P = P_1 \oplus P_2 \quad \Longleftrightarrow \quad P(x_2, x_3) = \sum_{\substack{x_2^1, x_3^1, x_2^2, x_3^2 \in \{0,e\} \\ x_2^1 + x_2^2 = x_2,\, x_3^1 + x_3^2 = x_3}} P_1(x_2^1, x_3^1) \cdot P_2(x_2^2, x_3^2) \quad \forall x_2, x_3 \in \{0, e\} \tag{53}$$

$$P = \alpha_1 P_1 + \alpha_2 P_2 \quad \Longleftrightarrow \quad P(x_2, x_3) = \alpha_1 P_1(x_2, x_3) + \alpha_2 P_2(x_2, x_3) \qquad \forall x_2, x_3 \in \{0, e\} \tag{54}$$

$$P = \Gamma(\overline{P}) \quad \Longleftrightarrow \quad P(x_2, x_3) = \sum_{\substack{\hat{x}_2, \bar{x}_3 \in \{0,e\}, \\ \bar{x}_3 \cdot (x_2 + \hat{x}_2) = x_3}} \overline{P}(x_2, \bar{x}_3) \cdot P_{\hat{\delta}_2}(\hat{x}_2) \qquad \forall x_2, x_3 \in \{0, e\} \tag{55}$$

value $\overline{P}_R^{(\ell)}$, which corresponds to a pair $(\mathrm{r}_{i,j}^{(2,\ell)}, \overline{\mathrm{r}}_{i,j}^{(3,\ell)})$ where $\overline{\mathrm{r}}_{i,j}^{(3,\ell)}$ is defined by,

$$\overline{\mathrm{r}}_{ij}^{(3,\ell)} = \begin{cases} y_{3,i}, & \ell = 0, \\ y_{3,i} \cdot \prod_{j' \in \mathcal{N}(i) \setminus \{j\}} \mathrm{l}_{j'i}^{(3,\ell)}, & \ell > 0. \end{cases}$$

That is, $\overline{\mathrm{r}}_{i,j}^{(3,\ell)}$ coincides with (24), except that the multiplication by $\hat{\mathrm{r}}_{ij}^{(2,\ell)}$ is omitted.

The weighted sums by λ_i, ρ_j and $\tilde{\lambda}_i$ in (48), (49) and (50) respectively, follow from the random construction of the computation graph, and are justified by the same arguments as [49, Expression (8)].

APPENDIX IV
PROOF OF LEMMA 1

We begin by writing,

$$I(\hat{\mathbf{Y}}_2^{\text{BP}}; \mathbf{Y}_2^{\text{BP}} \mid \mathbf{Y}_3) = H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3) - H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_2^{\text{BP}}, \mathbf{Y}_3) \tag{56}$$

Focusing on the first term on the right hand side of (56),

$$\begin{aligned} H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3) &\overset{(a)}{=} H(\hat{\mathbf{Y}}_2^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}} \mid \mathbf{Y}_3) \\ &= H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) + H(\hat{\mathbf{E}}_2^{\text{BP}} \mid \mathbf{Y}_3) \\ &\overset{(b)}{=} H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) + H(\hat{\mathbf{E}}_2^{\text{BP}}) \end{aligned} \tag{57}$$

where in (a), we have defined $\hat{\mathbf{E}}_2^{\text{BP}}$ as a random vector whose components are derived from $\hat{\mathbf{Y}}_2^{\text{BP}}$,

$$\hat{E}_{2,i}^{\text{BP}} \triangleq \begin{cases} e, & \hat{Y}_{2,i}^{\text{BP}} = e \\ 0, & \hat{Y}_{2,i}^{\text{BP}} \neq e. \end{cases} \tag{58}$$

To justify (b), we argue that $\hat{\mathbf{E}}_2^{\text{BP}}$ is independent of $\mathbf{Y}_3$. To see this, first observe that by the above definitions and (20), the following holds for $i = 1, ..n$,

$$\hat{E}_{2,i}^{\text{BP}} = E_{2,i}^{\text{BP}} + \hat{E}_{2,i} \tag{59}$$

where $E_{2,i}^{\text{BP}}$ is derived from $Y_{2,i}^{\text{BP}}$ in the same way as $\hat{E}_{2,i}^{\text{BP}}$ was derived from $\hat{Y}_{2,i}^{\text{BP}}$, and $\mathrm{E}_{2,i}$ is simply the erasure noise over the channel, defined as (12). By this definition, $\mathbf{E}_2^{\text{BP}}$ specifies the set of erased indices at the output of the relay's BP, and is thus a function of $\mathbf{E}_2$, the erasure noise on the channel from the source to the relay. Both $\mathbf{E}_2$ and $\hat{\mathbf{E}}_2$ are independent of $\mathbf{Y}_3$, and equality (b) now follows.

We now bound the first term on the right hand side of (57).

$$\begin{aligned} H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) &\leq \sum_{i=1}^{n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) \\ &\leq \sum_{i=1}^{n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{E}_{2,i}^{\text{BP}}, Y_{3,i}) \\ &\overset{(a)}{=} \sum_{i=1}^{n} H(X_{1,i} \mid \hat{E}_{2,i}^{\text{BP}} = 0, Y_{3,i}) \cdot \Pr[\hat{E}_{2,i}^{\text{BP}} = 0] \\ &\overset{(b)}{=} \sum_{i=1}^{n} H(X_{1,i} \mid \hat{E}_{2,i}^{\text{BP}} = 0, Y_{3,i} = e) \cdot \Pr[Y_{3,i} = e] \times \\ &\qquad \times \Pr[\hat{E}_{2,i}^{\text{BP}} = 0] \\ &\overset{(c)}{\leq} \sum_{i=1}^{n} 1 \cdot \delta_3 \cdot (1 - \eta_{2,i}^{\text{BP}} \circ \hat{\delta}_2) \\ &= n \left\{ \delta_3 \cdot \left[1 - \left(\frac{1}{n} \sum_{i=1}^{n} \eta_{2,i}^{\text{BP}} \right) \circ \hat{\delta}_2 \right] \right\} \\ &\overset{(d)}{=} n \Big\{ \delta_3 \cdot [1 - \delta_2^{\text{BP}} \circ \hat{\delta}_2] + o(1) \Big\} \end{aligned} \tag{60}$$

In (a), we have relied on the fact that if $\hat{E}_{2,i}^{\text{BP}} = e$, then by (58), $\hat{Y}_{2,i}^{\text{BP}} = e$ with probability 1 and so $H(\hat{Y}_{2,i}^{\text{BP}} | \hat{E}_{2,i}^{\text{BP}} = e, Y_{3,i}) = 0$. If $\hat{E}_{2,i}^{\text{BP}} = 0$ then $\hat{E}_{2,i}^{\text{BP}} = X_{1,i}$, where $X_{1,i}$ is a component of the transmitted source vector, $\mathbf{X}_1$. In (b), we have observed that if $Y_{3,i} = x$ where $x \in \{0, 1\}$, then $X_{1,i} = x$ with probability 1, and so $H(X_{1,i} \mid \hat{E}_{2,i}^{\text{BP}} = 0, Y_{3,i} = x) = 0$. In (c), we have observed that since $X_{1,i}$ is defined over $\{0, 1\}$, $H(X_{1,i} \mid \hat{E}_{2,i}^{\text{BP}} = 0, Y_{3,i} = e) \leq 1$. We have also evaluated $\Pr[Y_{3,i} = e] = \delta_3$. Finally, we defined $\eta_{2,i}^{\text{BP}} = \Pr[E_{2,i}^{\text{BP}} = e]$. By (59), and the fact of $\hat{E}_{2,i}$ is distributed as Erasure($\hat{\delta}_2$) (see Sec. III-C), invoking (5), we have $\Pr[\hat{E}_{2,i}^{\text{BP}} = e] = \eta_{2,i}^{\text{BP}} \circ \hat{\delta}_2$. Observe that each $\eta_{2,i}^{\text{BP}}$ is in fact a function of the code $\mathcal{C}$, which was randomly selected from the (λ, ρ) ensemble. Thus, it is a random variable. In (d), we have relied on the following

derivation,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\eta_{2,i}^{\text{BP}} &\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}\Pr[E_{2,i}^{\text{BP}}=e]\\ &\overset{(b)}{=} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\chi_{E_{2,i}^{\text{BP}}=e}\right]\\ &\overset{(c)}{=} \mathbb{E}\left[D_2^{\text{BP}}\right]\\ &\overset{(d)}{=} \delta_2^{\text{BP}}+o(1)\end{aligned} \tag{61}$$

In (a) we have invoked the definition of $\hat{\eta}_{2,i}^{\text{BP}}$. In (b), $\chi_{\hat{E}_{2,i}^{\text{BP}}=e}$ is an indicator random variable, which equals 1 if $\hat{E}_{2,i}^{\text{BP}}=e$. The expectation is over the channel transitions, but the code $\mathcal{C}$ is assumed to be fixed. In (c), D_2^{BP} is the realized erasure rate at the output of BP at the relay (see Definition 2). In (d) we have relied on Corollary 1 (see Appendix IV-A below). This bound holds for large enough n with probability at least $1-\exp(-\beta/2\cdot n^{1/3})$ for $\beta>0$, thus complying with the conditions of Lemma 1.

We now turn to bound the second term on the right hand side of (57).

$$\begin{aligned}H(\hat{\mathbf{E}}_2^{\text{BP}}) &\leq \sum_{i=1}^{n}H(\hat{E}_{2,i}^{\text{BP}})\\ &\overset{(a)}{=} \sum_{i=1}^{n}h(\eta_{2,i}^{\text{BP}}\circ\hat{\delta}_2)\\ &\overset{(b)}{\leq} n\cdot h\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{2,i}^{\text{BP}}\right)\\ &\overset{(c)}{=} n\cdot\left[h(\delta_2^{\text{BP}}\circ\hat{\delta}_2)+o(1)\right]\end{aligned} \tag{62}$$

In (a) $\eta_{2,i}^{\text{BP}}$ is defined as above. In (b) we have applied Jensen's inequality, relying on the concavity of the entropy function. In (c) we have relied on (61) and invoked the continuity of $h(\cdot)$.

We now turn to evaluate the second term on the right hand side of (56).

$$\begin{aligned}&H(\hat{\mathbf{Y}}_2^{\text{BP}}\,|\,\mathbf{Y}_2^{\text{BP}},\mathbf{Y}_3)=\\ &\quad\overset{(a)}{=} H(\hat{\mathbf{Y}}_2^{\text{BP}}\,|\,\mathbf{Y}_2^{\text{BP}})\\ &\quad\overset{(b)}{=} \sum_{i=1}^{n}H(\hat{Y}_{2,i}^{\text{BP}}\,|\,\mathbf{Y}_2^{\text{BP}},\hat{Y}_{2,1}^{\text{BP}},...,\hat{Y}_{2,i-1}^{\text{BP}})\\ &\quad\overset{(c)}{=} \sum_{i=1}^{n}H(\hat{Y}_{2,i}^{\text{BP}}\,|\,Y_{2,i}^{\text{BP}})\\ &\quad\overset{(d)}{=} \sum_{i=1}^{n}h(\hat{\delta}_2)(1-\eta_{2,i}^{\text{BP}})\\ &\quad= n\cdot\left[h(\hat{\delta}_2)\left(1-\frac{1}{n}\sum_{i=1}^{n}\eta_{2,i}^{\text{BP}}\right)\right]\\ &\quad\overset{(e)}{=} n\cdot\left[h(\hat{\delta}_2)\cdot(1-\delta_2^{\text{BP}})+o(1)\right]\end{aligned} \tag{63}$$

where in (a) we have relied on the fact that the three random vectors on both sides of the equation form a Markov chain: $\mathbf{Y}_3\leftrightarrow\mathbf{Y}_2^{\text{BP}}\leftrightarrow\hat{\mathbf{Y}}_2^{\text{BP}}$. In (b), we have simply applied the chain rule for entropy. In (c), we have relied on the fact that the random variables on the previous line make the following Markov chain: $\hat{\mathbf{Y}}_{2,\sim i}^{\text{BP}}\leftrightarrow\mathbf{Y}_{2,\sim i}^{\text{BP}}\leftrightarrow Y_{2,i}^{\text{BP}}\leftrightarrow\hat{Y}_{2,i}^{\text{BP}}$, where $\hat{\mathbf{Y}}_{2,\sim i}^{\text{BP}}$ and $\mathbf{Y}_{2,\sim i}^{\text{BP}}$ are defined as (1). In (d), we have relied on the observation that if $Y_{2,i}^{\text{BP}}=e$, then $\hat{Y}_{2,i}^{\text{BP}}=e$ with probability 1 and thus $H(\hat{Y}_{2,i}^{\text{BP}}\,|\,Y_{2,i}^{\text{BP}}=e)=0$, and if $Y_{2,i}^{\text{BP}}=x\in\{0,1\}$ then $\hat{Y}_{2,i}^{\text{BP}}=x$ with probability $1-\hat{\delta}_2$ and $\hat{Y}_{2,i}^{\text{BP}}=e$ with probability $\hat{\delta}_2$. We have also defined $\eta_{2,i}^{\text{BP}}$ as above. By definition, $E_{2,i}^{\text{BP}}=e$ if and only if $Y_{2,i}^{\text{BP}}=e$ and so $\eta_{2,i}^{\text{BP}}=\Pr[Y_{2,i}^{\text{BP}}=e]$. In (e) we have applied (61).

Finally, combining (56), (57), (60), (62) and (63), we obtain our desired (26).

□

A. Analysis of D_2^{BP}

Lemma 3: Let $\mathcal{C}$ and δ_2^{BP} be defined as in Lemma 1. Let D_2^{BP} be the realized erasure rate in an application of BP at the relay (see Definition 2). Then the following holds for large enough n (the probability being over the random selection of $\mathcal{C}$), with probability at least $1-\exp(-\beta/2\cdot n^{1/3})$,

$$\Pr\left[|D_2^{\text{BP}}-\delta_2^{\text{BP}}|>n^{-1/3}\,|\,\text{The code }\mathcal{C}\text{ is used}\right]\leq 2\mathrm{e}^{-\beta/2\cdot n^{1/3}} \tag{64}$$

where $\beta>0$ is some constant, dependent on λ,ρ and t.

Proof: By [47, Theorem 2], there exist constants $\beta,\gamma>0$, which are dependent on λ,ρ,t (where t is the number of BP iterations performed), such that for all $\epsilon>0$ and integer $n>0$ satisfying $n>2\gamma/\epsilon$,

$$\Pr\left[|D_2^{\text{BP}}-\delta_2^{\text{BP}}|>\epsilon\right] \leq 2\mathrm{e}^{-\beta\epsilon^2 n} \tag{65}$$

Letting $\epsilon=n^{-1/3}$ again we obtain that for all $n>(2\gamma)^{3/2}$,

$$\Pr\left[|D_2^{\text{BP}}-\delta_2^{\text{BP}}|>n^{-1/3}\right] \leq 2\mathrm{e}^{-\beta\cdot n^{1/3}} \tag{66}$$

The random space in (65) and (66) is comprised of the random channel transitions as well as the random selection of the code from the (λ,ρ) ensemble. For a fixed code $\mathcal{C}$ let $P(\mathcal{C})$ denote the left hand side of (66) conditioned on the use of $\mathcal{C}$. That is,

$$P(\mathcal{C})=\Pr\left[|D_2^{\text{BP}}-\delta_2^{\text{BP}}|>n^{-1/3}\,|\,\text{The code }\mathcal{C}\text{ is used}\right]$$

The random space in $P(\mathcal{C})$ consists of the channel transitions only. $P(\mathcal{C})$ itself is a random variable which depends on the randomly selected $\mathcal{C}$. We now use Markov's inequality to bound the probability that $P(\mathcal{C})$ is very large.

$$\begin{aligned}\Pr\left[P(\mathcal{C})>2\mathrm{e}^{-\beta/2\cdot n^{1/3}}\right] &\overset{(a)}{\leq} \frac{\mathbb{E}\left[P(\mathcal{C})\right]}{2\mathrm{e}^{-\beta/2\cdot n^{1/3}}}\\ &\overset{(b)}{=} \frac{\Pr\left[|D_2^{\text{BP}}-\delta_2^{\text{BP}}|>n^{-1/3}\right]}{2\mathrm{e}^{-\beta/2\cdot n^{1/3}}}\\ &\overset{(c)}{\leq} \mathrm{e}^{-\beta/2\cdot n^{1/3}}\end{aligned}$$

(a) follows by Markov's inequality. The expectation on the right hand side is over the random selection of the code $\mathcal{C}$. (b) follows by the law of total probability and the definition of $P(\mathcal{C})$ and (c) follows by (66). The result now follows. □

Corollary 1: Let $\mathcal{C},\delta_2^{\text{BP}}$ and D_2^{BP} be defined as in Lemma 3 and let $\mathbb{E}[D_2^{\text{BP}}]$ denote the expected value of D_2^{BP}, for a fixed

value of $\mathcal{C}$. Then the following holds for large enough n with probability at least $1-\exp(-\beta/2\cdot n^{1/3})$,

$$\Pr\left[\left|\mathbb{E}[D_2^{\text{BP}}]-\delta_2^{\text{BP}}\right|>n^{-1/3}+2\mathrm{e}^{-\beta/2\cdot n^{1/3}}\right]\leq \mathrm{e}^{-\beta/2\cdot n^{1/3}}$$

where $\beta>0$ is defined as in Lemma 3.

The corollary follows immediately from Lemma 3 by the observation that D_2^{BP} is confined to $[0,1]$.

APPENDIX V
PROOF OF THEOREM 4

Our proof is based on the proof of Lemma 1 (Appendix IV, above). As noted in Sec. III-F, our bound improves upon the bound of Lemma 1 by exploiting dependencies between the components of $\mathbf{Y}_2^{\text{BP}}$, the output of BP at the relay with soft-DF-BP. Specifically, in Appendix V-A below we will exploit dependencies between bits at discovered by BP (see Sec. III-F) to tighten the bound (60) on $H(\hat{\mathbf{Y}}_2^{\text{BP}}\mid\hat{\mathbf{E}}_2^{\text{BP}},\mathbf{Y}_3)$. In Appendix V-B we exploit dependencies between erasures at the output of BP, to produce a bound on $H(\hat{\mathbf{E}}_2^{\text{BP}})$ which is sometimes tighter than (62).

The proof of the theorem will be obtained by applying the bounds in Appendices V-A and V-B in the following way. $1/n\cdot I(\mathbf{Y}_2^{\text{BP}};\hat{\mathbf{Y}}_2^{\text{BP}}\mid\mathbf{Y}_3)$ is upper bounded by $I_1^+(\hat{\delta}_2)+o(1)$ by (56), (57), (62) and (63) from the proof of Lemma 1 and (69) (Appendix V-A below). $1/n\cdot I(\mathbf{Y}_2^{\text{BP}};\hat{\mathbf{Y}}_2^{\text{BP}}\mid\mathbf{Y}_3)$ is upper bounded by $I_2^+(\hat{\delta}_2)+o(1)$ by a similar set equations, replacing (62) with (78) (Appendix V-B below). Finally, the l.d.f. operation in (28) is justified in Appendix V-D.

A. Upper Bound on $H(\hat{\mathbf{Y}}_2^{BP}\mid\hat{\mathbf{E}}_2^{BP},\mathbf{Y}_3)$

As noted above, our bound in this section applies dependencies that exist between bits that were discovered by BP. An outline of the main idea behind the proof is provided in Sec. III-F.

We begin with the string of equations ending with (67) on the following page. In (a) $\mathbf{E}_2$ is defined as $\hat{\mathbf{E}}_2^{\text{BP}}$ was defined in (58), replacing $\hat{\mathbf{Y}}_2^{\text{BP}}$ with $\mathbf{Y}_2$. In this equation, we have relied on the Markov chain relation between the random variables, $\mathbf{Y}_3\leftrightarrow\hat{\mathbf{Y}}_2^{\text{BP}}\leftrightarrow\hat{\mathbf{E}}_2^{\text{BP}}\leftrightarrow\mathbf{E}_2$. In (b) we have applied the definition of conditional entropy: The expectation is over the variable $\boldsymbol{\mathcal{E}_2}$, which is defined to be distributed identically as $\mathbf{E}_2$. In (c) we have applied the chain rule for entropy. We assume, without loss of generality, that the components of $\hat{\mathbf{Y}}_2^{\text{BP}}$ are ordered in the order that they would have been discovered by BP in its equivalent simplified formulation, Algorithm 4 (Sec. III-F) had the channel erasures corresponded to $\boldsymbol{\mathcal{E}_2}$. That is, the first components are the ones revealed at iteration 0 of the algorithm (i.e., not erased by the channel), they are followed by the components that the algorithm revealed at iteration 1, and so forth. Components that were not revealed at any iteration are ordered last. In (d) we have separated the sums of components that were revealed by the channel, components that were revealed at iterations 1 and above of Simplified BP, and components not revealed by Simplified BP. $\mathcal{D}_2$ denotes the erasure rate of $\mathcal{E}_2$ (Definition 2), and $\mathcal{D}_2^{\text{BP}}$ denotes the erasure rate of $\mathbf{Y}_2^{\text{BP}}$ at the output of BP, and is distributed as D_2^{BP} (see Appendix IV-A).

We now examine the three sums on the right hand side of (67). The desired exploitation of the dependencies between the bits discovered by BP will take place in the second sum, which we will examine last. The third sum is easily evaluated to equal zero. This is because $\hat{Y}_{2,i}^{\text{BP}}=e$ with probability 1 for all components of the sum, which follows from (20) because $Y_{2,i}^{\text{BP}}=e$ at these components. Turning to the components of the first sum, let $i\in\{1,...,(1-\mathcal{D}_2)n\}$.

$$\begin{aligned}
&H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{Y}_{2,1}^{\text{BP}},...,\hat{Y}_{2,i-1}^{\text{BP}},\hat{\mathbf{E}}_2^{\text{BP}},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2},\mathbf{Y}_3)\leq\\
&\quad\overset{(a)}{\leq}H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{E}_{2,i}^{\text{BP}},Y_{3,i},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2})\\
&\quad=H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{E}_{2,i}^{\text{BP}}=e,Y_{3,i})\Pr[\hat{E}_{2,i}^{\text{BP}}=e\mid\mathbf{E}_2=\boldsymbol{\mathcal{E}_2}]+\\
&\qquad+H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{E}_{2,i}^{\text{BP}}=0,Y_{3,i})\Pr[\hat{E}_{2,i}^{\text{BP}}=0\mid\mathbf{E}_2=\boldsymbol{\mathcal{E}_2}]\\
&\quad\overset{(b)}{=}H(X_{1,i}\mid Y_{3,i})\cdot(1-\hat{\delta}_2)\\
&\quad\overset{(c)}{\leq}\delta_3(1-\hat{\delta}_2)
\end{aligned}\tag{68}$$

In (a), we have reduced the conditions on the entropy to obtain an upper bound on its values. In (b) we have applied $H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{E}_{2,i}^{\text{BP}}=e,Y_{3,i})=0$, which holds because conditioned on $\hat{E}_{2,i}^{\text{BP}}=e$ we have that $\hat{Y}_{2,i}^{\text{BP}}=e$ with probability 1. We have also relied on the fact that if $\hat{E}_{2,i}^{\text{BP}}=0$ then $\hat{Y}_{2,i}^{\text{BP}}=X_{1,i}$, where $X_{1,i}$ is the transmitted signal from the source at time i. Finally, we are currently examining $i\in\{1,...,(1-\mathcal{D}_2)n\}$, for which $\mathcal{E}_{2,i}=0$ by definition. If $E_{2,i}=\mathcal{E}_{2,i}=0$ we clearly also have $E_{2,i}^{\text{BP}}=0$. By (59) we have $\Pr[\hat{E}_{2,i}^{\text{BP}}=0\mid\mathbf{E}_2=\boldsymbol{\mathcal{E}_2}]=\Pr[\hat{E}_{2,i}=0]=(1-\hat{\delta}_2)$. In (c), we have relied on the fact that $H(X_{1,i}\mid Y_{3,i}=e)=H(X_{1,i})\leq 1$ and $H(X_{1,i}\mid Y_{3,i}=0)=H(X_{1,i}\mid Y_{3,i}=1)=0$.

We now turn to the components of the second sum in (67). Let $i\in\{(1-\mathcal{D}_2)n+1,...,(1-\mathcal{D}_2^{\text{BP}})n\}$.

$$\begin{aligned}
&H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{Y}_{2,1}^{\text{BP}},...,\hat{Y}_{2,i-1}^{\text{BP}},\hat{\mathbf{E}}_2^{\text{BP}},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2},\mathbf{Y}_3)\leq\\
&\quad\overset{(a)}{\leq}H(\hat{Y}_{2,i}^{\text{BP}}\mid\hat{Y}_{2,j_1}^{\text{BP}},...,\hat{Y}_{2,j_{d-1}}^{\text{BP}},\hat{E}_{2,i}^{\text{BP}},Y_{3,i},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2})\\
&\quad\overset{(b)}{\leq}H(\hat{Y}_{2,i}^{\text{BP}}\mid\overline{Y}_{2,i}^{\text{BP}},\hat{E}_{2,i}^{\text{BP}},Y_{3,i},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2})\\
&\quad=H(X_{1,i}\mid\overline{Y}_{2,i}^{\text{BP}},Y_{3,i},\mathbf{E}_2=\boldsymbol{\mathcal{E}_2})\cdot(1-\hat{\delta}_2)\\
&\quad\overset{(c)}{\leq}\delta_3\left(1-(1-\hat{\delta}_2)^{d-1}\right)(1-\hat{\delta}_2)
\end{aligned}$$

The analysis follows in the line of the derivation leading to (68), and we will elaborate only on the differences. Recall that each component at indices $i\in\{(1-\mathcal{D}_2)n+1,...,(1-\mathcal{D}_2^{\text{BP}})n\}$ was discovered in the application of Simplified BP. Let $j_1,...,j_{d-1}$ be the indices of the other variable nodes that were connected to a check node by which index i was discovered. By nature of the Simplified BP algorithm, these indices necessarily correspond to bits that were discovered at previous iterations of Simplified BP. Thus, since we have assumed that the indices $i=1,...,n$ are arranged by the order in which components of $\mathbf{Y}_2^{\text{BP}}$ were discovered by Simplified BP, we have $\{j_1,...,j_{d-1}\}\subset\{1,...,i-1\}$. (a) now follows.

In (b), we have defined $\overline{Y}_{2,i}^{\text{BP}}=\hat{Y}_{2,j_1}^{\text{BP}}+...+\hat{Y}_{2,j_{d-1}}^{\text{BP}}$, where addition is modulo-2. In (c), we have relied on the fact that $\overline{Y}_{2,i}^{\text{BP}}$ is erasure if any of $\hat{Y}_{2,j_1}^{\text{BP}},...,\hat{Y}_{2,j_{d-1}}^{\text{BP}}$ is erasure. Conditioned on $\mathbf{E}_2=\boldsymbol{\mathcal{E}_2}$, each of these variables is erasure with probability $\hat{\delta}_2$ (this follows as in our above analysis

$$
\begin{aligned}
H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) &\overset{(a)}{=} H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2, \mathbf{Y}_3) \\
&\overset{(b)}{=} \mathbb{E}_{\boldsymbol{\mathcal{E}_2}} \Big[H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}, \mathbf{Y}_3) \Big] \\
&\overset{(c)}{=} \mathbb{E}_{\boldsymbol{\mathcal{E}_2}} \left[\sum_{i=1}^{n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{Y}_{2,1}^{\text{BP}}, ..., \hat{Y}_{2,i-1}^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}, \mathbf{Y}_3) \right] \\
&\overset{(d)}{=} \mathbb{E}_{\boldsymbol{\mathcal{E}_2}} \Bigg[\sum_{i=1}^{(1-\mathcal{D}_2)n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{Y}_{2,1}^{\text{BP}}, ..., \hat{Y}_{2,i-1}^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}, \mathbf{Y}_3) + \\
&\qquad + \sum_{i=(1-\mathcal{D}_2)n+1}^{(1-\mathcal{D}_2^{\text{BP}})n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{Y}_{2,1}^{\text{BP}}, ..., \hat{Y}_{2,i-1}^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}, \mathbf{Y}_3) + \\
&\qquad + \sum_{i=(1-\mathcal{D}_2^{\text{BP}})n+1}^{n} H(\hat{Y}_{2,i}^{\text{BP}} \mid \hat{Y}_{2,1}^{\text{BP}}, ..., \hat{Y}_{2,i-1}^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}, \mathbf{Y}_3) \Bigg] \qquad (67)
\end{aligned}
$$

of $\Pr[\hat{E}_{2,i}^{\text{BP}} = 0 \mid \mathbf{E}_2 = \boldsymbol{\mathcal{E}_2}]$) and thus $\Pr[\overline{Y}_{2,i}^{\text{BP}} = e] = 1 - (1 - \hat{\delta}_2)^{d-1}$.

We now return to (67). Relying on the above discussion, we now have the string of equations ending with (69) on the following page. In (a), $\mathcal{D}_2$ is the erasure rate of the channel output $\mathbf{Y}_2$ at the relay. Its expected value is clearly δ_2. We have also relied on Corollary 1 (Appendix IV-A above) to express the expected values of $\mathcal{D}_2^{\text{BP}}$, recalling that it is identically distributed as D_2^{BP}.

This bound (69) concludes our analysis of $H(\hat{\mathbf{Y}}_2^{\text{BP}} | \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3)$.

□

B. Upper Bound on $H(\hat{\mathbf{E}}_2^{BP})$

As noted above, our bound in this section applies dependencies that exist between erasures at the output of BP. An outline of the main idea behind the proof is again provided in Sec. III-F. We begin as follows.

$$
\begin{aligned}
H(\hat{\mathbf{E}}_2^{\text{BP}}) &\leq H(\mathbf{E}_2^{\text{BP}}, \hat{\mathbf{E}}_2^{\text{BP}}) \\
&= H(\mathbf{E}_2^{\text{BP}}) + H(\hat{\mathbf{E}}_2^{\text{BP}} \mid \mathbf{E}_2^{\text{BP}}) \qquad (70)
\end{aligned}
$$

where $\mathbf{E}_2^{\text{BP}}$ is as defined in Appendix IV, following (59). Focusing on the second term on the right hand side of (70) we obtain,

$$
\begin{aligned}
H(\hat{\mathbf{E}}_2^{\text{BP}} \mid \mathbf{E}_2^{\text{BP}}) &\leq \sum_{i=1}^{n} H(\hat{E}_{2,i}^{\text{BP}} \mid E_{2,i}^{\text{BP}}) \\
&\overset{(a)}{=} \sum_{i=1}^{n} h(\hat{\delta}_2)(1 - \eta_{2,i}^{\text{BP}}) \\
&\overset{(b)}{=} n \Big[(1 - \delta_2^{\text{BP}}) \cdot h(\hat{\delta}_2) + o(1) \Big] \qquad (71)
\end{aligned}
$$

In (a), we have observed that if $E_{2,i}^{\text{BP}} = e$, then by (59), $\hat{E}_{2,i}^{\text{BP}} = e$ with probability 1 and thus $H(\hat{E}_{2,i}^{\text{BP}} \mid E_{2,i}^{\text{BP}} = e) = 0$. If $E_{2,i}^{\text{BP}} = 0$ then $\hat{E}_{2,i}^{\text{BP}} = \hat{E}_2$ and thus $H(\hat{E}_{2,i}^{\text{BP}} \mid E_{2,i}^{\text{BP}} = 0) = h(\hat{\delta}_2)$. We have also defined $\eta_{2,i}^{\text{BP}}$ to equal $\Pr[E_{2,i}^{\text{BP}} = e]$ as in Appendix IV. Finally, (b) follows in the same lines as in our derivation of (63) in Appendix IV.

We now turn to the first term in (70).

$$
\begin{aligned}
H(\mathbf{E}_2^{\text{BP}}) &\overset{(a)}{\leq} H(\mathbf{E}_2^{\text{BP}}, D_2^{\text{BP}}) \\
&= H(\mathbf{E}_2^{\text{BP}} \mid D_2^{\text{BP}}) + H(D_2^{\text{BP}}) \\
&\overset{(a)}{\leq} H(\mathbf{E}_2^{\text{BP}} \mid D_2^{\text{BP}}) + \log(n+1) \qquad (72)
\end{aligned}
$$

where (a) we have defined D_2^{BP} to equal the erasure rate of $\mathbf{E}_2^{\text{BP}}$ (see Definition 2). In (b) we have relied on the fact that D_2^{BP} is confined to the set $\{0, 1/n, 2/n, ..., 1\}$, which contains $n+1$ elements.

The vector $\mathbf{E}_2^{\text{BP}}$ specifies the bits that remained undecoded at the output of BP. Di *et al.* [14, Lemma 1.1] proved that these bits correspond to a *stopping set* of the code $\mathcal{C}$ (see [14] for its definition). For $s \in (0, n)$, let $N_{\mathcal{C}}(s)$ denote the number of stopping sets of size s in $\mathcal{C}$. We thus have,

$$H(\mathbf{E}_2^{\text{BP}} \mid D_2^{\text{BP}} = \alpha) \leq \log N_{\mathcal{C}}(\alpha n)$$

Plugging this into (72) we obtain the string of equations ending with (73) on the following page. Note that the expectations in these equations are over D_2^{BP}. We begin by examining the second additive term in (73).

$$
\begin{aligned}
\mathbb{E}\Big[\log N_{\mathcal{C}}(D_2^{\text{BP}} n) \,\Big|\, |D_2^{\text{BP}} - \delta_2^{\text{BP}}| > n^{-1/3} \Big] &\times \\
\times \Pr\Big[|D_2^{\text{BP}} - \delta_2^{\text{BP}}| > n^{-1/3} \Big] & \\
\leq \log(2^n) \cdot 2\mathrm{e}^{-\beta/2 \cdot n^{1/3}} &= o(1) \qquad (74)
\end{aligned}
$$

where we have relied on the fact that the number $N_{\mathcal{C}}(D_2^{\text{BP}} n)$ of stopping sets of size $D_2^{\text{BP}} n$ is trivially less than 2^n, which is the number of subsets of of the indices $\{1, ..., n\}$. We have also applied Lemma 3 (Appendix IV-A).

We now turn to the first additive term in (73). Burshtein and Miller [7, Theorem 9] and Orlitsky *et al.* [43, Theorem 5] examined $\mathbb{E}[N_{\mathcal{C}}(\alpha n)]$ where the expectation is over all codes $\mathcal{C}$ in the (λ, d) LDPC ensemble. From their development we have, for all $\alpha = k/n, \quad k = 0, ..., n$,

$$\frac{1}{n} \log \mathbb{E}\Big[N_{\mathcal{C}}(\alpha n) \Big] \leq f(\alpha) + o(1) \qquad (75)$$

$$\begin{aligned} H(\hat{\mathbf{Y}}_2^{\text{BP}} \mid \hat{\mathbf{E}}_2^{\text{BP}}, \mathbf{Y}_3) &\leq \mathbb{E}_{\boldsymbol{\mathcal{E}_2}}\left[n\mathcal{D}_2 \cdot \delta_3(1-\hat{\delta}_2) + n(\mathcal{D}_2 - \mathcal{D}_2^{\text{BP}})\delta_3\left(1-(1-\hat{\delta}_2)^{d-1}\right)(1-\hat{\delta}_2)\right] \\ &\overset{(a)}{=} n\left[\delta_2 \cdot \delta_3(1-\hat{\delta}_2) + (\delta_2 - \delta_2^{\text{BP}})\delta_3\left(1-(1-\hat{\delta}_2)^{d-1}\right)(1-\hat{\delta}_2) + o(1)\right] \\ &= n\cdot\delta_3(1-\hat{\delta}_2)\left[(1-\delta_2) + (\delta_2-\delta_2^{\text{BP}})\left(1-(1-\hat{\delta}_2)^{d-1}\right) + o(1)\right] \end{aligned} \tag{69}$$

$$\begin{aligned} H(\mathbf{E}_2^{\text{BP}}) &\leq \mathbb{E}\Big[\ \log N_{\mathcal{C}}(D_2^{\text{BP}} n)\ \Big] + \log(n+1) \\ &= \mathbb{E}\left[\ \log N_{\mathcal{C}}(D_2^{\text{BP}} n)\ \middle|\ |D_2^{\text{BP}} - \delta_2^{\text{BP}}| \leq n^{-1/3}\right]\cdot \Pr\left[|D_2^{\text{BP}} - \delta_2^{\text{BP}}| \leq n^{-1/3}\right] + \\ &\quad \mathbb{E}\left[\ \log N_{\mathcal{C}}(D_2^{\text{BP}} n)\ \middle|\ |D_2^{\text{BP}} - \delta_2^{\text{BP}}| > n^{-1/3}\right]\cdot \Pr\left[|D_2^{\text{BP}} - \delta_2^{\text{BP}}| > n^{-1/3}\right] + \log(n+1) \end{aligned} \tag{73}$$

where $f(\alpha)$ is given by (32) and the term $o(1)$ is independent of α. A few minor remarks are deferred to Appendix V-C below.

In (75), the expectation is over all codes $\mathcal{C}$ in our ensemble. In our analysis, however, we are interested in the probability that an individual code $\mathcal{C}$ has $N_{\mathcal{C}}(\alpha n)$ that greatly exceeds $f(\alpha)$. As in the proof of Lemma 3 (Appendix IV-A), we apply Markov's inequality to bound this probability. For fixed n we let $\hat{f}(\alpha; n)$ denote the left hand side of (75). We now derive,

$$\begin{aligned} \Pr\left[N_{\mathcal{C}}(\alpha n) > \mathrm{e}^{n(\hat{f}(\alpha;n)+n^{-1/2})}\right] &\leq \frac{\mathbb{E}[N_{\mathcal{C}}(\alpha n)]}{\mathrm{e}^{n(\hat{f}(\alpha;n)+n^{-1/2})}} \\ &= \frac{\mathrm{e}^{n\hat{f}(\alpha;n)}}{\mathrm{e}^{n(\hat{f}(\alpha;n)+n^{-1/2})}} \ = \ \mathrm{e}^{-n^{1/2}} \end{aligned}$$

where the probability is over the random selection of a code $\mathcal{C}$ from the (λ, d) ensemble. By a union bound we obtain,

$$\begin{aligned} \Pr\left[\exists \alpha \in \{0, 1/n, 2/n, ..., 1\}\ :\ N_{\mathcal{C}}(\alpha n) > \mathrm{e}^{n(\hat{f}(\alpha;n)+n^{-1/2})}\right] \\ \leq (n+1)\cdot \mathrm{e}^{-n^{1/2}} \end{aligned}$$

By these results, for large enough n, with probability at least $1-\exp(-n^{1/3})$ (as required by Theorem 4's conditions), a randomly selected code $\mathcal{C}$ satisfies,

$$\frac{1}{n}\log N_{\mathcal{C}}(\alpha n) \leq f(\alpha) + o(1) \quad \forall \alpha \in \{0, 1/n, 2/n, ..., 1\} \tag{76}$$

In the remainder of the proof below, will assume that our $\mathcal{C}$ satisfies (76).

Combining (73), (74) and (76) we obtain,

$$\begin{aligned} H(\mathbf{E}_2^{\text{BP}}) &\leq n\left\{\mathbb{E}\left[\ f(D_2^{\text{BP}})\ \middle|\ |D_2^{\text{BP}} - \delta_2^{\text{BP}}| \leq n^{-1/3}\right] + o(1)\right\} \\ &= n\Big[f(\delta_2^{\text{BP}}) + o(1)\Big] \end{aligned} \tag{77}$$

where we have invoked the continuity of $f(\alpha)$ which holds by [43, Corollary 6]. Finally, combining (70), (71) and (77) we obtain our desired bound on $H(\hat{\mathbf{E}}_2^{\text{BP}})$.

$$H(\hat{\mathbf{E}}_2^{\text{BP}}) \leq n\Big[f(\delta_2^{\text{BP}}) + (1-\delta_2^{\text{BP}})\cdot h(\hat{\delta}_2) + o(1)\Big] \tag{78}$$

□

C. Some Remarks Regarding Equation (75)

Our expression (32) for $f(\alpha)$ is a slight variation of [43, Theorem 5] ($\gamma(\alpha)$ in their notation). In [43], expressions for the minimizers x and y of the various minimizations (denoted x_0 and y_0) are provided, and the expression for $f(\alpha)$ is provided as a function of them. The range of the maximization of β is also different from the one we used in (32). An examination of their proof shows that these differences do not affect the final outcome.

We now discuss (75) (most importantly, with the $o(1)$ term being independent of α). To justify its validity, we argue that in [43, Theorem 5], adding the term $1/n\log n$ to the right hand side of the equation, produces an upper bound on $1/n\log\mathbb{E}[N_C(\alpha n)]$ for all n. To see this, observe that in [43, Lemmas 3 and 4] each limit may be replaced by a supremum over all n. This holds by replacing the asymptotic saddle-point analysis in the lemmas' proofs with an upper bound as [7, Equation (6)]. In [43, Equation (11)], where these lemmas were applied, we may discard the limit, replace the sum by a supremum, and add a compensation term $1/n\log n$, to obtain a bound on $1/n\log\mathbb{E}[N_C(\alpha n)]$ rather than an evaluation of its limit. The desired result will then follow as in the proof of [43, Theorem 5].

D. Justification of the l.d.f. operator

The operator l.d.f. in (28) is easily justified by the fact that $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$ *must* be descending as a function of $\hat{\delta}_2$. To see this, let $\hat{\delta}_2' < \hat{\delta}_2''$ and let $\hat{\mathbf{Y}}_2^{\text{BP}'}$ and $\hat{\mathbf{Y}}_2^{\text{BP}''}$ be random vectors whose components are defined based on (20),

$$\begin{aligned} \hat{Y}_{2,i}^{\text{BP}'} &= Y_{2,i}^{\text{BP}} + \hat{\mathbf{E}}_{2,i}' \\ \hat{Y}_{2,i}^{\text{BP}''} &= Y_{2,i}^{\text{BP}} + \hat{\mathbf{E}}_{2,i}'' \end{aligned}$$

where the components $\hat{E}_{2,i}'$ and $\hat{E}_{2,i}''$ are independent and distributed as Erasure($\hat{\delta}_2'$) and Erasure($\hat{\delta}_2''$), respectively. Then $\hat{\mathbf{Y}}_2^{\text{BP}''}$ is stochastically degraded with respect to $\hat{\mathbf{Y}}_2^{\text{BP}'}$ and thus $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}''} \mid \mathbf{Y}_3) < I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}'} \mid \mathbf{Y}_3)$, implying the desired monotonicity of $I(\mathbf{Y}_2^{\text{BP}}; \hat{\mathbf{Y}}_2^{\text{BP}} \mid \mathbf{Y}_3)$ as a function of $\hat{\delta}_2$.

APPENDIX VI
PROOFS FOR SEC. III-G

A. Proof of Theorem 5

We begin by introducing the following notation, for an arbitrary code $\mathcal{C}$ and $\delta \in [0,1]$,

$$I(\mathcal{C};\ \delta) \triangleq I(\mathbf{X};\ \mathbf{Y})$$

where $\mathbf{X}$ is uniformly distributed within the codewords of $\mathcal{C}$ and $\mathbf{Y}$ is randomly related to $\mathbf{X}$ via the transition probabilities of a BEC(δ). By Fano's inequality (e.g. [13, Sec. 8.9]) and by virtue of the "goodness' of the sequence $\mathcal{C}_n$, the following must hold at $\delta = \delta^\star$:

$$\lim_{n\to\infty} \frac{1}{n} I(\mathcal{C}_n;\ \delta^\star) = R = 1 - \delta^\star \tag{79}$$

The following lemma relates $I(\mathcal{C}_n;\ \delta)$ to our desired $P_{\text{MAP}}(\mathcal{C}_n;\ \delta)$, and parallels [23, Expression (1)]. The lemma extends results from [44].

Lemma 4: The following holds for any linear code $\mathcal{C}$ and $\delta \in [0,1]$,

$$\frac{d}{d\delta} I(\mathcal{C};\ \delta) = n \cdot \left(-\frac{1}{\delta}\right) \cdot P_{\text{MAP}}(\mathcal{C};\ \delta) \tag{80}$$

Proof: We begin with the following identity, which follows from [44, Expression (8)] (similar expressions are available in [39, Theorem 1] and [2, Theorem 1])

$$\frac{d}{d\delta} I(\mathcal{C};\ \delta) = \\ = \sum_{i=1}^{n} \mathbb{E}\left[\frac{\partial \ln P^{\delta}_{Y_i|X_i}(Y_i|X_i)}{\partial \delta} \log P^{\delta}_{X_i|\mathbf{Y}}(X_i|\mathbf{Y})\right] \tag{81}$$

The expectation is over both $\mathbf{X}$ and $\mathbf{Y}$. $P^{\delta}_{Y_i|X_i}(y|x)$ and $P^{\delta}_{X_i|\mathbf{Y}}(x|\mathbf{y})$ denote the conditional probability functions corresponding to $\mathbf{X}$ and $\mathbf{Y}$, where the superscript δ denotes the BEC erasure probability.

Rewriting (81) we obtain,

$$\frac{d}{d\delta} I(\mathcal{C};\ \delta) = \mathbb{E}_{\mathbf{Y}}\left\{\sum_{i=1}^{n} \mathbb{E}_{X_i}\left[\frac{\partial \ln P^{\delta}_{Y_i|X_i}(Y_i|X_i)}{\partial \delta} \times \right.\right. \\ \left.\left. \times \log P^{\delta}_{X_i|\mathbf{Y}}(X_i|\mathbf{Y})\right]\right\} \tag{82}$$

where the first expectation is over $\mathbf{Y}$ and the second over X_i.

Let $\hat{X}_i(\mathbf{y})$ denote the MAP decoder output corresponding to a channel output vector $\mathbf{y}$ and index i. As mentioned in Sec. III-G, this output is obtained by mapping of the *a posteriori* probability $P^{\delta}_{X_i|\mathbf{Y}}(1|\mathbf{y})$ to the set $\{0,1,e\}$,

$$\hat{X}_i(\mathbf{y}) = \begin{cases} x, & P^{\delta}_{X_i|\mathbf{Y}}(1|\mathbf{y}) = x \in \{0,1\}; \\ e, & P^{\delta}_{X_i|\mathbf{Y}}(1|\mathbf{y}) = 1/2. \end{cases}$$

where we have relied on the fact that since $\mathcal{C}$ is linear, $P^{\delta}_{X_i|\mathbf{Y}}(1|\mathbf{y})$ is guaranteed to be in the set $\{0,1,1/2\}$ [48, Sec. 3.2.1].

If $\hat{X}_i(\mathbf{y}) = x \in \{0,1\}$, the transmitted X_i, conditioned on $\mathbf{Y} = \mathbf{y}$, equals x with probability 1, and thus $\log P^{\delta}_{X_i|\mathbf{Y}}(X_i|\mathbf{y}) = 0$ with probability 1. If $\hat{X}_i(\mathbf{y}) = e$ we have $P^{\delta}_{X_i|\mathbf{Y}}(x_i|\mathbf{y}) = 1/2$ for $x_i \in \{0,1\}$. Furthermore, $\mathbf{y}$ must clearly satisfy $y_i = e$ (or else $\hat{X}_i(\mathbf{y}) = e$ cannot hold) and thus $P^{\delta}_{Y_i|X_i}(y_i \mid x_i) = \delta$ for $x_i \in \{0,1\}$. We can now rewrite (82) as,

$$\begin{aligned} \frac{d}{d\delta} I(\mathcal{C};\ \delta) &= \mathbb{E}_{\mathbf{Y}}\left\{\sum_{\hat{X}_i(\mathbf{Y})=e} \mathbb{E}_{X_i}\left[\frac{\partial \ln \delta}{\partial \delta} \log(1/2)\right]\right\} \\ &= \cdot \mathbb{E}_{\mathbf{Y}}\left\{\sum_{\hat{X}_i(\mathbf{Y})=e} \left(-\frac{1}{\delta}\right)\right\} \\ &= \left(-\frac{1}{\delta}\right) \cdot \mathbb{E}_{\mathbf{Y}}\left(\left|\left\{i : \hat{X}_i(\mathbf{Y}) = e\right\}\right|\right) \end{aligned}$$

The desired (80) now follows by the definition of $P_{\text{MAP}}(\mathcal{C};\ \delta)$. □

We now prove (34) for $\delta > \delta^\star$. Let $\epsilon > 0$ and assume that $P_{\text{MAP}}(\mathcal{C}; \delta_o) < \delta_o - \epsilon$ for some $\delta_o > \delta^\star + \epsilon$. By (80) we have,

$$\begin{aligned} \frac{1}{n} I(\mathcal{C};\ \delta^\star) &= \frac{1}{n} I(\mathcal{C};\ 1) - \int_{\delta^\star}^{1} \left(-\frac{1}{\delta}\right) \cdot P_{\text{MAP}}(\mathcal{C};\ \delta)\, d\delta \\ &\overset{(a)}{=} \int_{\delta^\star}^{1} \frac{1}{\delta} \cdot P_{\text{MAP}}(\mathcal{C};\ \delta)\, d\delta \\ &\overset{(b)}{\leq} \int_{[\delta^\star, \delta_o-\epsilon]\cup[\delta_o,1]} \frac{1}{\delta} \cdot \delta\, d\delta + \\ &\qquad + \int_{[\delta_o-\epsilon,\delta_o]} \frac{1}{\delta} \cdot (\delta_o - \epsilon)\, d\delta \\ &= (1-\delta^\star) - \left[\epsilon - (\delta_o - \epsilon) \ln\left(\frac{\delta_o}{\delta_o - \epsilon}\right)\right] \\ &\overset{(c)}{=} (1-\delta^\star) - h(\delta_o, \epsilon) \end{aligned}$$

In (a), we have used the fact that $I(\mathcal{C};\ 1) = 0$ which can be verified straightforwardly. In (b) we have relied on the fact that the MAP decoder outputs no more erasures than it obtains via the channel output, and thus $P_{\text{MAP}}(\mathcal{C}; \delta) \leq \delta$. Also, to obtain that $P_{\text{MAP}}(\mathcal{C};\ \delta) < \delta_o - \epsilon$ for $\delta \in [\delta_o - \epsilon, \delta_o]$ we have relied on our assumption $P_{\text{MAP}}(\mathcal{C};\ \delta_o) < \delta_o - \epsilon$ and on the fact that $P_{\text{MAP}}(\mathcal{C}; \delta)$ is non-descending as a function of δ. This holds because if $\delta_1 < \delta_2$, then BEC(δ_2) is stochastically degraded with respect to BEC(δ_1). In (c) we have simply defined $h(\delta_o, \epsilon)$ to equal the content of the brackets in the preceding equality.

$h(\delta_o, \epsilon)$ clearly approaches zero as $\epsilon \to 0$. It is also strictly positive for all $\epsilon > 0$. This follows from $\ln(\delta_o/(\delta_o - \epsilon)) < \epsilon/(\delta_o - \epsilon)$ which holds by the well-known inequality $\ln(1 + x) < x$ for all $x \neq 0, x > -1$.

The desired result (34) at $\delta_o > \delta^\star$ now follows using the following argument: Let $\epsilon > 0$ be small enough such that $\delta^\star < \delta_o - \epsilon$. Then for large enough k, we must have $P_{\text{MAP}}(\mathcal{C}_n; \delta_o) \geq \delta_o - \epsilon$ or else $1/n \cdot I(\mathcal{C}_n;\ \delta^\star) \leq (1-\delta^\star) - h(\delta_o, \epsilon) < 1 - \delta^\star$, thus violating (79).

We now turn to prove (34) in the range $\delta < \delta^\star$. The proof is obtained straightforwardly by examining the output of ML decoding. A ML decoder can be perceived as a suboptimal bitwise estimator, which is not allowed to output erasures. The bit error rate (normalized by n) at the output of ML

decoding cannot exceed the word error rate (denoted $P_e(\mathcal{C}_n;\delta)$ in Definition 4), because the worst-case number of bit errors in a decoded codeword cannot exceed n. By the "goodness' of $\{\mathcal{C}_n\}$, $P_e(\mathcal{C}_n;\delta)$ must approach zero for $\delta<\delta^\star$. The bit error rate with optimal estimation equals half the erasure rate at the output of bitwise MAP estimation as defined in Sec. III-G (the optimal bitwise estimator makes a uniform random in $\{0,1\}$ whenever the MAP estimator of Sec. III-G outputs an erasure). This bit error cannot exceed $P_e(\mathcal{C}_n;\delta)$, and thus must approach zero as well.

□

B. Proof of Theorem 6

We now focus on fixed n. For simplicity of notation, we drop the index n in $\mathcal{C}_n$. We begin by examining the right hand side of (36). Applying the chain rule for mutual information,

$$I(\mathbf{Y}_2;\hat{\mathbf{Y}}_2\mid\mathbf{Y}_3)=I(\hat{\mathbf{Y}}_2;\mathbf{Y}_2,\mathbf{Y}_3)-I(\hat{\mathbf{Y}}_2;\mathbf{Y}_3) \tag{83}$$

We now examine the first term on the right hand side of (83).

$$\begin{aligned}I(\hat{\mathbf{Y}}_2;\mathbf{Y}_2,\mathbf{Y}_3)&=H(\hat{\mathbf{Y}}_2)-H(\hat{\mathbf{Y}}_2\mid\mathbf{Y}_2,\mathbf{Y}_3)\\&=\Big[H(\hat{\mathbf{Y}}_2\mid\mathbf{X}_1)+I(\mathbf{X}_1;\hat{\mathbf{Y}}_2)\Big]-H(\hat{\mathbf{Y}}_2\mid\mathbf{Y}_2)\end{aligned} \tag{84}$$

where $\mathbf{X}_1$ is the transmitted codeword, and is uniformly distributed in $\mathcal{C}$. The term $I(\mathbf{X}_1;\hat{\mathbf{Y}}_2)$ in (84) will cancel out later. The other two terms can be evaluated,

$$\begin{aligned}H(\hat{\mathbf{Y}}_2\mid\mathbf{X}_1)&\overset{(a)}{=}\sum_{i=1}^{n}H(\hat{Y}_{2,i}\mid\mathbf{X}_1,\hat{Y}_{2,1},...,\hat{Y}_{2,i-1})\\&\overset{(b)}{=}\sum_{i=1}^{n}H(\hat{Y}_{2,i}\mid X_{1,i})\\&\overset{(c)}{=}n\cdot h(\delta_2\circ\hat{\delta}_2)\end{aligned} \tag{85}$$

where (a) follows by the chain rule for entropy, (b) follows by the Markov relation $\hat{Y}_{2,i}\leftrightarrow X_{1,i}\leftrightarrow(\mathbf{X}_1,\hat{Y}_{2,1},...,\hat{Y}_{2,i-1})$. Finally, (c) follows by (12), (16) and (5). Similarly, we obtain:

$$H(\hat{\mathbf{Y}}_2\mid\mathbf{Y}_2)=n\cdot(1-\delta_2)h(\hat{\delta}_2) \tag{86}$$

We now proceed to the second term on on the right hand side of (83).

$$\begin{aligned}I(\hat{\mathbf{Y}}_2;\mathbf{Y}_3)&\overset{(a)}{=}I(\mathbf{Y}_3;\mathbf{X}_1,\hat{\mathbf{Y}}_2)-I(\mathbf{Y}_3;\mathbf{X}_1\mid\hat{\mathbf{Y}}_2)\\&\overset{(b)}{=}I(\mathbf{Y}_3;\mathbf{X}_1)-I(\mathbf{Y}_3;\mathbf{X}_1\mid\hat{\mathbf{Y}}_2)\\&\overset{(c)}{=}I(\mathbf{Y}_3;\mathbf{X}_1)-\Big[I(\mathbf{Y}_3,\hat{\mathbf{Y}}_2;\mathbf{X}_1)-I(\hat{\mathbf{Y}}_2;\mathbf{X}_1)\Big]\end{aligned} \tag{87}$$

(a) and (c) follow by the chain rule for mutual information, and (b) follows by the Markov relation $\mathbf{Y}_3\leftrightarrow\mathbf{X}_1\leftrightarrow\hat{\mathbf{Y}}_2$. We now argue that the following holds:

$$I(\mathbf{Y}_3;\mathbf{X}_1)=n\cdot[(1-\delta_3)+o(1)] \tag{88}$$

To show this, we invoke the analysis of Appendix VI-A. With the notation of that appendix, $I(\mathbf{Y}_3;\mathbf{X}_1)=n\cdot I(\mathcal{C};\,\delta_3)$. By Lemma 4, the following holds.

$$\begin{aligned}\frac{1}{n}I(\mathcal{C};\,\delta_3)&\overset{(a)}{=}\frac{1}{n}I(C;\,\delta^\star)+\int_{\delta^\star}^{\delta_3}\left(-\frac{1}{\delta}\right)P_{\text{MAP}}(\mathcal{C};\,\delta)\,d\delta\\&\overset{(b)}{\geq}\Big[(1-\delta^\star)+o(1)\Big]+(\delta_3-\delta^\star)\\&=(1-\delta_3)+o(1)\end{aligned} \tag{89}$$

In (a), $\delta^\star=1-R$ is the Shannon limit for rate R (Definition 1). We have relied on (35) and the fact that $\delta_2\circ\hat{\delta}_2\leq 1$ to deduce $\delta_3\geq\delta^\star$. In (b), we have invoked the "goodness" of the sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$ and (79) to obtain $1/n\cdot I(C;\,\delta^\star)=(1-\delta^\star)+o(1)$. We have also relied on $P_{\text{MAP}}(\mathcal{C};\delta)\leq\delta$ as explained in Appendix VI-A. Finally (88) is obtained from (89) by the observation that the capacity of a BEC(δ_3) is $1-\delta_3$ and thus $1/n\cdot I(\mathcal{C};\,\delta_3)\leq 1-\delta_3$.

Turning to $I(\mathbf{Y}_3,\hat{\mathbf{Y}}_2;\mathbf{X}_1)$, we begin by arguing that

$$I(\mathbf{Y}_3,\hat{\mathbf{Y}}_2;\mathbf{X}_1)=I(\mathbf{Z};\mathbf{X}_1), \tag{90}$$

where $\mathbf{Z}=\mathbf{Y}_3\cdot\hat{\mathbf{Y}}_2$ (multiplication in $\mathbf{Y}_3\cdot\hat{\mathbf{Y}}_2$ is defined as in Sec. II-C). $I(\mathbf{Y}_3,\hat{\mathbf{Y}}_2;\mathbf{X}_1)\geq I(\mathbf{Z};\mathbf{X}_1)$ holds straightforwardly by the data processing inequality. To see why the reverse inequality holds, we define vectors $\hat{\mathbf{Y}}_2'$ and $\mathbf{Y}_3'$ that are stochastic functions of $\mathbf{Z}$, such that the joint distribution of the pair $(\hat{\mathbf{Y}}_2',\mathbf{Y}_3')$ and $\mathbf{X}_1$ is identical to that of $(\hat{\mathbf{Y}}_2,\mathbf{Y}_3)$ and $\mathbf{X}_1$. The inequality will then again be obtained by the data processing inequality.

Recall that by (12), the components of $\mathbf{Y}_3$ are related to those of $\mathbf{X}_1$ via the memoryless BEC(δ_3). By (12) and (16), the components of $\hat{\mathbf{Y}}_2$ are related to those of $\mathbf{X}_1$ via the memoryless BEC($\delta_2\circ\hat{\delta}_2$). We define the components of $(\hat{\mathbf{Y}}_2',\mathbf{Y}_3')$ in the following way:

$$(\hat{Y}_{2,i}',Y_{3,i}')=\begin{cases}(e,e), & Z_i=e;\\(e,Z_i), & Z_i\neq e,\text{ with probability }\epsilon_1;\\(Z_i,e), & Z_i\neq e,\text{ with probability }\epsilon_2;\\(Z_i,Z_i), & Z_i\neq e,\text{ with probability }\epsilon_3.\end{cases}$$

where,

$$\epsilon_1=\frac{(\delta_2\circ\hat{\delta}_2)\cdot(1-\delta_3)}{1-(\delta_2\circ\hat{\delta}_2)\cdot\delta_3},\quad\epsilon_2=\frac{(1-\delta_2\circ\hat{\delta}_2)\cdot\delta_3}{1-(\delta_2\circ\hat{\delta}_2)\cdot\delta_3}$$
$$\epsilon_3=\frac{(1-\delta_2\circ\hat{\delta}_2)\cdot(1-\delta_3)}{1-(\delta_2\circ\hat{\delta}_2)\cdot\delta_3}$$

Simple arithmetic now reveals that $(\hat{Y}_{2,i}',Y_{3,i}')$, conditioned on the value of $X_{1,i}$, are distributed identically as $(\hat{Y}_{2,i},Y_{3,i})$.

The channel from $\mathbf{X}_1$ to $\mathbf{Z}$ is a memoryless BEC with crossover probability $(\delta_2\circ\hat{\delta}_2)\cdot\delta_3$. By arguments similar to those used to prove (86) we obtain,

$$I(\mathbf{Z};\mathbf{X}_1)=n\cdot\left[\Big(1-(\delta_2\circ\hat{\delta}_2)\cdot\delta_3\Big)+o(1)\right] \tag{91}$$

Combining (83), (84), (85), (86), (87) (88), (90) and (91), we obtain our desired (36).

□

C. Proof that $R_{UB} \leq R_{CF}$

We now argue that R_{UB} in (38) is upper bounded by R_{CF} in (19). Let R be contained in the set on the right hand side of (38), and let $\hat{\delta}_2$ be an accompanying value as specified there.

First assume $R \leq 1 - \delta_3$. A simple examination of the content of the braces in (19) reveals that it is at least $1 - \delta_3$ for all $\hat{\delta}_2$, and thus $R_{\text{CF}} \geq 1 - \delta_3 \geq R$.

Now assume $R > 1 - \delta_3$. We first show that we may assume, without loss of generality, that the condition $R \leq 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3$ in (38) is satisfied by equality. If this does not hold, then we can increase $\hat{\delta}_2$ until an equality is reached. Increasing $\hat{\delta}_2$ can only reduce the left hand side of (37) (see Appendix V-D above). Thus, increasing $\hat{\delta}_2$ does not violate (37). Finally, by Theorem 6, condition (37) implies condition (18) (this can be seen by taking n to infinity on the right hand side of (36)), and thus $R = 1 - (\delta_2 \circ \hat{\delta}_2) \cdot \delta_3 \leq R_{\text{CF}}$.

□

APPENDIX VII
PROOF OF THEOREM 7

In our analysis, we focus on the first source-destination pair. An outline of the proof was provided in Sec. IV-C. We let X_1^* and X_2^* denote scalar random variables that are distributed as in the discussion following (40) and (41). That is, both are uniformly distributed in $\{\pm 1\}$. We also let Y_1^* be a random variable that is related to them via the channel transition equation (39).

We distinguish between two cases, $R < I(X_2^*; Y_1^* \mid X_1^*)$ and $R \geq I(X_2^*; Y_1^* \mid X_1^*)$. In the first case, which is discussed in Appendix VII-A below, we prove that $R \leq R_{\text{MUD}}$. In the second case, which is discussed in Appendix VII-B, we prove $R \leq R_{\text{SUD}}$. The desired (42) thus follows.

A. Analysis in the Range $R < I(X_2^*; Y_1^* \mid X_1^*)$

Our proof begins in lines similar to the proof of the converse of the capacity of the multiple-access channel, [13, Sec. 14.3.4].

$$\begin{aligned}
n \cdot 2R &\overset{(a)}{=} n\Big(R_{1,n} + R_{2,n} + o(1)\Big) \\
&\overset{(b)}{=} H(W_1, W_2) + n \cdot o(1) \\
&= I(W_1, W_2; \mathbf{Y}_1) + H(W_1, W_2 \mid \mathbf{Y}_1) + n \cdot o(1) \\
&= I(W_1, W_2; \mathbf{Y}_1) + H(W_1 \mid \mathbf{Y}_1) + \\
&\qquad + H(W_2 \mid W_1, \mathbf{Y}_1) + n \cdot o(1) \\
&\overset{(c)}{=} I(W_1, W_2; \mathbf{Y}_1) + n \cdot o(1) + n \cdot o(1) + n \cdot o(1) \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{1i}) + n \cdot o(1) \\
&\overset{(e)}{\leq} nI(X_1^*, X_2^*; Y_1^*) + n \cdot o(1)
\end{aligned} \tag{92}$$

In (a), $R_{1,n}$ and $R_{2,n}$ are the rates of the codes $\mathcal{C}_{1,n}$ and $\mathcal{C}_{2,n}$, respectively, and the equality holds by the definition of R as the rate of the code sequences $\{\mathcal{C}_{1,n}\}_{n=1}^{\infty}$ and $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$. In (b), W_1 and W_2 are the messages that were transmitted by Sources 1 and 2, respectively, defined as [13, Sec. 14.3.4]. The equality holds because the two messages are statistically independent, and uniformly distributed in $\{1, .., 2^{nR_{1,n}}\}$ and $\{1, .., 2^{nR_{2,n}}\}$, respectively. In (c), $H(W_1 \mid \mathbf{Y}_1) = n \cdot o(1)$ holds by Fano's inequality [13, Theorem 2.11.1], relying on the fact that the probability of error in the decoding of W_1, by the conditions of our Theorem 7, approaches zero with n. The justification for $H(W_2 \mid W_1, \mathbf{Y}_1) = n \cdot o(1)$ will be provided shortly. (d) follows by the same arguments as [13, Equation (14.116)]. Finally, (e) follows by our discussion in Appendix VII-C below.

By (92), recalling that we are now focusing our attention to the range $R < I(X_2^*; Y_1^* \mid X_1^*)$, we obtain by (40) (recalling our above definitions of X_1^*, X_2^* and Y_1^*), $R \leq R_{\text{MUD}}$.

We now prove $H(W_2 \mid W_1, \mathbf{Y}_1) = n \cdot o(1)$ in the above equation (c). Consider the scenario facing a decoder of W_2 (at Destination 1), recalling the channel equation (39). As noted in Sec. IV-C, given W_1, the decoder is able to eliminate X_1, and is thus faced with a point-to-point BIAWGN communication scenario, with SNR $= h^2/\sigma^2$. The capacity of this channel is clearly $C(\text{SNR}) = I(X_2^*; Y_1^* \mid X_1^*)$. By the fact that $R < I(X_2^*; Y_1^* \mid X_1^*)$ (we are currently focusing on such R), we have that SNR $>$ SNR*, where SNR* is the Shannon limit for rate R. By the "goodness" of $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$, recalling Definition 4, we obtain that the probability of error, under maximum-likelihood decoding, of W_2 given $\mathbf{Y}_1$ and W_1, must approach zero with n. Thus, by Fano's inequality (as in our analysis of $H(W_1 | \mathbf{Y}_1)$), we obtain $H(W_2 | W_1, \mathbf{Y}_1) = n \cdot o(1)$ as desired.

B. Analysis in the Range $R \geq I(X_2^*; Y_1^* \mid X_1^*)$

Our analysis begins as in Appendix VII-A.

$$\begin{aligned}
n \cdot 2R &\overset{(a)}{=} I(W_1, W_2; \mathbf{Y}_1) + H(W_1 \mid \mathbf{Y}_1) + \\
&\qquad + H(W_2 \mid W_1, \mathbf{Y}_1) + n \cdot o(1) \\
&\overset{(b)}{\leq} I(W_1, W_2; \mathbf{Y}_1) + n \cdot o(1) + \\
&\qquad + n\Big(R - I(X_2^*; Y_1^* \mid X_1^*) + o(1)\Big) + n \cdot o(1) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{1i}) + nR - nI(X_2^*; Y_1^* \mid X_1^*) + \\
&\qquad + n \cdot o(1) \\
&\overset{(d)}{\leq} nI(X_1^*, X_2^*; Y_1^*) + nR - nI(X_2^*; Y_1^* \mid X_1^*) + n \cdot o(1) \\
&\overset{(e)}{=} nI(X_1^*; Y_1^*) + nR + n \cdot o(1) \\
&\overset{(f)}{=} nR_{\text{SUD}} + nR + n \cdot o(1)
\end{aligned} \tag{93}$$

(a) follows as in Appendix VII-A. In (b), $H(W_1 | \mathbf{Y}_1) = n \cdot o(1)$ follows again as in Appendix VII-A, and $H(W_2 \mid W_1, \mathbf{Y}_1) \leq n(R - I(X_2^*; Y_1^* \mid X_1^*) + o(1))$ will be justified shortly. (c) and (d) follow as in Appendix VII-A. (e) follows by the chain rule for mutual information [13, Theorem 2.5.2]. Finally, (f) follows by (41), recalling our above definitions of X_1^* and Y_1^*.

Subtracting nR from both sides of the above inequality, dividing by n and taking n to infinity, we obtain $R \leq R_{\text{SUD}}$ as desired.

We now prove $H(W_2 | W_1, \mathbf{Y}_1) \leq n(R - I(X_2^*; Y_1^* | X_1^*) +$

$o(1)$), as required in inequality (b) above.

$$\begin{aligned} H(W_2 \mid W_1, \mathbf{Y}_1) &\overset{(a)}{\leq} H(\mathbf{X}_2 \mid W_1, \mathbf{Y}_1) + n \cdot o(1) \\ &= H(\mathbf{X}_2 \mid W_1) - I(\mathbf{X}_2; \mathbf{Y}_1 \mid W_1) + n \cdot o(1) \\ &\overset{(b)}{\leq} n\Big(R + o(1)\Big) - I(\mathbf{X}_2; \mathbf{Y}_1 \mid W_1) + n \cdot o(1) \\ &\overset{(c)}{=} nR - I(\mathbf{X}_2; \tilde{\mathbf{Y}}_2) + n \cdot o(1) \\ &\overset{(d)}{\leq} nR - n\Big(C(\text{SNR}) + o(1)\Big) + n \cdot o(1) \\ &\overset{(e)}{=} nR - nI(X_2^*; Y_1^* \mid X_1^*) + n \cdot o(1) \end{aligned} \tag{94}$$

(a) is proven in Appendix VII-D. (b) follows by the fact that the cardinality of the range of the random vector $\mathbf{X}_2$ cannot exceed that of W_2, which is $2^{nR_{2,n}}$, and $R_{2,n} = R + o(1)$. In (c), we have defined $\tilde{\mathbf{Y}}_2 \triangleq \mathbf{Y}_1 - \mathbf{X}_1(W_1)$, where $\mathbf{X}_1(W_1)$ is the codeword corresponding to W_1. As noted in Sec. IV-C and Appendix VII-A, by (39), the channel from $\mathbf{X}_2$ to $\tilde{\mathbf{Y}}_2$ is a BIAWGN with $\text{SNR} = h^2/\sigma^2$. The capacity of this channel is clearly $C(\text{SNR}) = I(X_2^*; Y_1^* \mid X_1^*)$. As we have confined our attention to $R \geq I(X_2^*; Y_1^* | X_1^*)$, we have $\text{SNR} \leq \text{SNR}^\star$ where $\text{SNR}^\star$ is again the Shannon limit for rate R. The justification for (d) will be provided shortly. Finally, in (e) we have simply rewritten $C(\text{SNR}) = I(X_2^*; Y_1^* \mid X_1^*)$.

To justify (d), we argue that $I(\mathbf{X}_2; \tilde{\mathbf{Y}}_2) \geq n(C(\text{SNR}) + o(1))$. Had the SNR satisfied $\text{SNR} = \text{SNR}^\star$, this would have held trivially by the "goodness" of code sequence $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$ and Fano's inequality. However, as mentioned above, we are now interested in $\text{SNR} \leq \text{SNR}^\star$. Our justification in this range of SNR follows by similar arguments to those leading to (89). Specifically, we let $I(\mathcal{C}; \text{SNR}) = I(\mathbf{X}; \mathbf{Y})$ where $\mathbf{X}$ is uniformly distributed within the code $\mathcal{C}$ and $\mathbf{Y}$ is related to it via the transitions of a BIAWGN, as (2). With this definition, by our above discussion $I(\mathbf{X}_2; \tilde{\mathbf{Y}}_2) = I(\mathcal{C}_{2,n}, \text{SNR})$. We let $C(\text{SNR}^\star)$ denote the capacity of a BIAWGN with an SNR of $\text{SNR}^\star$. We now have,

$$\begin{aligned} nC(\text{SNR}^\star) &\overset{(a)}{=} I(\mathcal{C}_{2,n}; \text{SNR}^\star) + n \cdot o(1) \\ &\overset{(b)}{=} I(\mathcal{C}_{2,n}; \text{SNR}) + \int_{\text{SNR}}^{\text{SNR}^\star} \frac{1}{2}\text{mmse}(\mathcal{C}_{2,n}; \text{snr}) d\text{snr} + n \cdot o(1) \\ &\overset{(c)}{\leq} I(\mathcal{C}_{2,n}; \text{SNR}) + \int_{\text{SNR}}^{\text{SNR}^\star} \frac{1}{2} n \cdot \text{mmse}(\text{bitwise}; \text{snr}) d\text{snr} + \\ &\quad + n \cdot o(1) \\ &\overset{(d)}{=} I(\mathcal{C}_{2,n}; \text{SNR}) + n \cdot \Big(C(\text{SNR}^\star) - C(\text{SNR})\Big) + n \cdot o(1) \end{aligned} \tag{95}$$

(a) follows by similar arguments to (79), relying on Fano's inequality and the "goodness" of $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$. In (b), the mmse function is defined as (45) (Appendix I). The equality follows from the relation between mutual information and the MMSE, see [23, Equation (1)]. In (c), $\text{mmse}(\text{bitwise}; \text{snr})$ denotes the MMSE in the estimation of a symbol X which is uniformly distributed in $\{\pm 1\}$, from Y which is related to X via a BIAWGN with noise variance $1/\text{snr}$. In such estimation, the decoder does not have the benefit of the code structure to draw upon, and so the estimation error clearly increases in comparison to the estimation of a given bit in $\mathcal{C}_{2,n}$. In (d), we have relied on the fact that the derivative of the function $C(\text{SNR})$ with respect to SNR is $1/2\text{mmse}(\text{bitwise}; \text{snr})$. This follows from the discussion of [23, Sec. II.A][27]. Finally, recalling $I(\mathbf{X}_2; \tilde{\mathbf{Y}}_2) = I(\mathcal{C}_{2,n}, \text{SNR})$, we have our desired result.

□

C. Analysis of $I(X_{1i}, X_{2i}; Y_{1i})$

We now prove inequality (e) in the string of equations leading to (92) and inequality (d) in the string of equations leading to (93). Our proof relies on the properties of "good" codes for the BIAWGN. Specifically, we show that the marginal distributions of the individual code symbols X_{1i} and X_{2i}, $i = 1, ..., n$, cannot stray too far from the uniform distribution over $\{\pm 1\}$, which is the capacity-achieving distribution over the BIAWGN [21, Theorem 4.5.1]. Our proof is a variation of a similar result by [54, Theorem 4].

$$\begin{aligned} \sum_{i=1}^{n} I(X_{1i}, X_{2i}; Y_{1i}) &\overset{(a)}{=} \sum_{i=1}^{n} i_2\Big(p_{1i} \times p_{2i}\Big) \\ &\overset{(b)}{\leq} n \cdot i_2\Big(\frac{1}{n}\sum_{i=1}^{n}(p_{1i} \times p_{2i})\Big) \\ &\overset{(c)}{=} n \cdot i_2\Big(p^* \times p^* + o(1)\Big) \\ &\overset{(d)}{=} n \cdot \Big[i_2\Big(p^* \times p^*\Big) + o(1)\Big] \\ &\overset{(e)}{=} n \cdot I(X_1^*, X_2^*; Y_1^*) + n \cdot o(1) \end{aligned} \tag{96}$$

In (a), we have made the following definitions. $i_2(\cdot)$ is a function whose argument is a probability function $p(x_1, x_2)$ where $(x_1, x_2) \in \{0, 1\}^2$. Its value is $I(X_1, X_2; Y_1)$ where (X_1, X_2) are distributed as $p(x_1, x_2)$ and Y_1 is related to them via the transitions of the interference channel, (39). p_{1i} is a probability function over $x_1 \in \{0, 1\}$, corresponding to the distribution of X_{1i}. p_{2i} is similarly defined, corresponding to X_{2i}. $p_{1i} \times p_{2i}$ is defined by,

$$p = p_{1i} \times p_{2i} \quad \Rightarrow \quad p(x_1, x_2) = p_{1i}(x_1) \cdot p_{2i}(x_2) \quad \forall (x_1, x_2) \in \{0, 1\}^2$$

The independence between X_{1i} and X_{2i}, implied by equality (a), follows from the independence between the messages W_1 and W_2, as in [13, Equation (14.122)].

Inequality (b) follows by Jensen's inequality and the concavity of the mutual information as a function of the marginals of its distributions, [13, Theorem 2.7.4]. In (c), we have defined p^* to be the distribution of X_1^* (and of X_2^*). The justification for this equality will be provided shortly. (d) follows by the continuity of $i_2(\cdot)$, and (e) follows by its above definition.

To prove equality (c) above, we consider communication over a point-to-point BIAWGN with an SNR equal to the Shannon limit for rate R, $\text{SNR}^\star$. We let $i_1(p)$ denote $I(X; \hat{Y})$ where X takes the value 1 with probability p and -1 with probability $1 - p$. $\hat{Y}$ is related to X via the transitions of the

[27] Specifically, [23, Equation (17)] corresponds to $\text{mmse}(\text{bitwise}; \text{snr})$ and [23, Equation (18)] corresponds to $C(\text{SNR})$.

above-mentioned BIAWGN, see (2).

$$\begin{aligned} n \cdot i_1\left(\frac{1}{2}\right) &\overset{(a)}{\leq} I(\mathbf{X}_1;\hat{\mathbf{Y}}_1) + n\cdot o(1) \\ &\overset{(b)}{\leq} \sum_{i=1}^{n} I(X_{1i};\hat{Y}_{1i}) + n\cdot o(1) \\ &\overset{(c)}{=} \sum_{i=1}^{n} i_1(\pi_{1i}) + n\cdot o(1) \\ &\overset{(d)}{=} \sum_{i=1}^{n} i_1(\hat{\pi}_{1i}) + n\cdot o(1) \\ &\overset{(e)}{\leq} n\cdot i_1\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\pi}_{1i}\right) + n\cdot o(1) \\ &\overset{(f)}{\leq} n\cdot i_1\left(\frac{1}{2}\right) + n\cdot o(1) \end{aligned}$$

In (a), $\hat{\mathbf{Y}}_1$ corresponds to the output of the above mentioned BIAWGN channel, when provided with $\mathbf{X}_1$ as its input. We have relied on the fact that as the capacity-achieving distribution for the BIAWGN corresponds to $p = 1/2$, the capacity of the BIAWGN is $i_1(1/2)$. The equality now follows by the same arguments as equality (a) in the string of equations ending with (95), relying on the "goodness" of $\{\mathcal{C}_{1,n}\}_{n=1}^{\infty}$. (b) follows as [13, Equation (8.104)], relying on the memorylessness of the BIAWGN. In (c), we have defined π_{1i} to be the probability that X_{1i} is equal to 1. In (d), we have defined $\hat{\pi}_{1i} = \min(\hat{\pi}_{1i}, 1-\pi_{1i})$ and the equality follows by the obvious symmetry of $i_1(\cdot)$. In (e), we have again applied Jensen's inequality and [13, Theorem 2.7.4]. In (f), we have relied on the fact that the maximum of $i_1(\cdot)$ is achieved at $1/2$, as this corresponds to the capacity-achieving distribution of the BIAWGN.

We now have,

$$\lim_{n\to\infty} i_1\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\pi}_{1i}\right) = i_1\left(\frac{1}{2}\right)$$

The function $i_1(\cdot)$ achieves its maximum uniquely at $p = 1/2$. We thus obtain,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\hat{\pi}_{1i} = \frac{1}{2} \tag{97}$$

We now define,

$$f_1(n) \triangleq \frac{1}{2} - \frac{1}{n}\sum_{i=1}^{n}\hat{\pi}_{1i} \tag{98}$$

and,

$$I_1(n) = \left\{ i \ : \ \hat{\pi}_{1i} \geq \frac{1}{2} - \sqrt{f_1(n)} \right\}$$

By a simple argument, relying on (98) and the fact that $\hat{\pi}_{1i} \leq 1/2$ for all i, we have,

$$|I_1(n)^c| \leq \sqrt{f_1(n)}\cdot n$$

where $I_1(n)^c$ denotes the complement set of $I_1(n)$. Note that $I_1(n)$ satisfies,

$$\begin{aligned} &i \in I_1(n) \Rightarrow \\ &\quad \frac{1}{2} - \sqrt{f_1(n)} \leq p_{1i}(x) \leq \frac{1}{2} + \sqrt{f_1(n)} \qquad \forall x \in \{\pm 1\} \end{aligned}$$

where $p_{1i}(\cdot)$ is as defined above. We similarly define $f_2(n)$ and $I_2(n)$. Equation (96) now follows by the observation that $f_1(n)$ and $f_2(n)$ approach zero with n, and by the above definition of p^*.

D. Analysis of $H(W_2 \mid W_1, \mathbf{Y}_1)$

We now justify inequality (a) in the string of equations leading to (94).

$$\begin{aligned} H(W_2 \mid W_1, \mathbf{Y}_1) &\leq H(W_2, \mathbf{X}_2 \mid W_1, \mathbf{Y}_1) \\ &= H(\mathbf{X}_2 \mid W_1, \mathbf{Y}_1) + H(W_2 \mid \mathbf{X}_2, W_1, \mathbf{Y}_1) \\ &\overset{(a)}{=} H(\mathbf{X}_2 \mid W_1, \mathbf{Y}_1) + H(W_2 \mid \mathbf{X}_2) \\ &\overset{(b)}{\leq} H(\mathbf{X}_2 \mid W_1, \mathbf{Y}_1) + H(W_2 \mid \hat{\mathbf{Y}}) \\ &\overset{(c)}{=} H(\mathbf{X}_2 \mid W_1, \mathbf{Y}_1) + n\cdot o(1) \end{aligned}$$

(a) follows by the Markov chain relation $W_2 \leftrightarrow \mathbf{X}_2 \leftrightarrow (W_1, \mathbf{Y}_1)$. In (b), we have defined $\hat{\mathbf{Y}}$ to be the output of a BIAWGN channel (2), whose SNR is equal to SNR*, which is provided with the input $\mathbf{X}_2$. The inequality follows by the data processing inequality, using the Markov chain relation $W_2 \leftrightarrow \mathbf{X}_2 \leftrightarrow \hat{\mathbf{Y}}$. By the "goodness" of $\{\mathcal{C}_{2,n}\}_{n=1}^{\infty}$, recalling Definition 4, the probability of error, when decoding W_2 from $\hat{\mathbf{Y}}$, must approach zero with n. Equality (c) now follows, using Fano's inequality.

APPENDIX VIII
RESULTS FOR SEC. V

A. Optimization of Codes for Erasure Relay Channels

We now elaborate our algorithm for the design of LDPC codes for erasure relays channels, as discussed in Sec. V-A. The input to the algorithm is a pair (λ, ρ). $\hat{\delta}_2$ is obtained from (λ, ρ) by selecting the minimum value such that $I^+(\hat{\delta}_2)$ as defined (28) is less than or equal to C_o. ρ is kept constant throughout the iterations of the algorithm, and λ is iteratively improved. In our description below, we let λ denote the left edge distribution at the beginning of an iteration, and λ^+ the improved distribution obtained at the iteration's end. We also let $\hat{\delta}_2^+$ be obtained from (λ^+, ρ) in the same way as $\hat{\delta}_2$ was obtained from (λ, ρ). The algorithm seeks to maximize the design rate corresponding to (λ^+, ρ) while requiring that λ^+ be "close enough" to λ (as explained below) so that $(\lambda^+, \rho, \hat{\delta}_2^+)$ will likely still be admissible (see Sec. V-A). The algorithm stops when a *non*-admissible triplet is obtained.

We define "closeness" by,

$$\begin{aligned} &\left| \sum_i \lambda_i^+ \cdot P_R^{(\ell,i)}(x_2, x_3) - P_R^{(\ell)}(x_2, x_3) \right| \leq \\ &\qquad \eta \cdot \left| P_R^{(\ell-1)}(x_2, x_3) - P_R^{(\ell)}(x_2, x_3) \right| \\ &\qquad\qquad \forall x_2, x_3 \in \{0, e\}, \ell = 1, ..., t \end{aligned} \tag{99}$$

where $P_R^{(\ell-1)}$, $P_R^{(\ell)}$ and $P_R^{(\ell,i)}$ are computed by an application of sim-DE corresponding to $(\lambda, \rho, \hat{\delta}_2)$. $P_R^{(\ell-1)}$ and $P_R^{(\ell)}$ denote rightbound message distributions (see Sec. III-E). Each $P_R^{(\ell,i)}$ is a *singleton* distribution, which equals the contents of the brackets in (48) and is computed as a byproduct of sim-DE. $\eta > 0$ is a design parameter (we experimented with $\eta = 0.1$). The constraints (99) are linear[28] in λ_i^+. We augment them with the requirement that the components of λ^+ sum to 1 and be confined to the range $[0, 1]$. By (8), maximization of the design rate is equivalent to maximizing $\sum_i \lambda_i^+ / i$. Thus, the maximization problem is linear, and can be solved by a linear program.

B. Optimization of Codes for Symmetric BIAWGN Interference Channels

The optimization algorithm proceeds in lines similar to those of Appendix VIII-A. Optimization begins with an *admissible* (λ, ρ) pair. We define an admissible pair as such that with its use, the bit error rate, at the output of each destination decoder, with respect to the codeword transmitted by the corresponding source (but not the interfering codeword), and as computed by density evolution, is sufficiently low (typically we require a BER of at most 10^{-5}). Optimization proceeds by attempting to iteratively improve λ, so that at each iteration, the design rate is increased, without violating the admissibility of (λ, ρ). To achieve this, we limit our search range to λ^+ that are "close" to λ, as defined below, where λ and λ^+ are defined as in Appendix VIII-A.

We define "closeness" by four sets of inequalities. The first set is given by,

$$\left| \sum_i \lambda_i^+ \cdot s_1^{(\ell,i)} - s_1^{(\ell)} \right| \leq \eta \cdot \left| s_1^{(\ell-1)} - s_1^{(\ell)} \right|, \quad \forall \ell = 1, ..., t$$

$\eta > 0$ is a fixed constant. $s_1^{(\ell,i)}$ and $s_1^{(\ell)}$ are computed based on an application of density evolution with (λ, ρ). $s_1^{(\ell)}$ is obtained from the rightbound message distribution $P_1^{(\ell)}$ with the primary decoder (Definition 6), at iteration ℓ, by $s_1^{(\ell)} = \Phi(P_1^{(\ell)})$, where,

$$\Phi(P) \triangleq \mathbb{E}\left[\log\left(1 + \mathrm{e}^{-L}\right)\right]$$

where the L is a random variable distributed as P. Each $s_1^{(\ell,i)}$ is similarly obtained from the *singleton* distribution $P_1^{(\ell,i)}$, defined as the intermediate distribution computed at iteration ℓ of density evolution, corresponding to rightbound messages computed at variable nodes of degree i at iteration ℓ of soft-IC-BP.

The second set of inequalities resembles the first, with λ^+, $s_1^{(\ell,i)}$ and $s_1^{(\ell)}$ replaced by $\tilde{\lambda}^+$, $\tilde{s}_1^{(\ell,i)}$ and $\tilde{s}_1^{(\ell)}$. $\tilde{\lambda}^+$ is obtained from λ^+ as (7). $\tilde{s}_1^{(\ell,i)}$ and $\tilde{s}_1^{(\ell)}$ are defined as $s_1^{(\ell,i)}$ and $s_1^{(\ell)}$, except that they are based on the distributions of messages from variable to state nodes, as computed by density evolution. Note that by multiplying both sides of the inequalities by $\sum_k (\lambda_k^+ / k)$, we again obtain inequalities that are linear in λ^+. The third and fourth sets of inequalities are the same as the above two, but replacing $s_1^{(\ell,i)}, s_1^{(\ell)}, \tilde{s}_1^{(\ell,i)}, \tilde{s}_1^{(\ell)}$ with $s_2^{(\ell,i)}, s_2^{(\ell)}, \tilde{s}_2^{(\ell,i)}, \tilde{s}_2^{(\ell)}$, computed based on distributions corresponding to the *interference* decoder (Definition 6).

We have also found it useful to further restrict λ^+ by the following constraint, which is reminiscent of the stability condition [49].

$$\lambda_2^+ \cdot \sum_j (j-1)\rho_j \leq (1 - \eta') \cdot \mathbb{E}\left[\mathrm{e}^{-L/2}\right]$$

where $\eta' > 0$ is a small constant (we experimented with $\eta' = 0.02$). L is a random variable, distributed as $P_1^{(t,2)}$, the singleton distribution corresponding to the last iteration of the primary decoder. As in Appendix VIII-A, we further augment these inequalities by requiring the components λ^+ to sum to 1, and be confined to the range $[0, 1]$. By seeking to maximize $\sum_i \lambda_i^+ / i$, we again obtain a linear maximization problem, which can be solved by linear programming.

The above procedure is prone to attraction to local maxima as follows. In our application of the procedure, we typically select the initial admissible pair (λ, ρ) to have very low design rate (even negative-valued). At such low rates, it is usually possible to achieve complete (rather than partial) decoding of both the primary and interference codewords. As the iterations of the optimization procedure progress, the design rate of the pair (λ, ρ) increases. However, the procedure remains confined to pairs with which complete decoding of both codewords is possible. The maximum possible rate in such conditions is upper bounded by R_{MUD} (see (40)).

This problem is easily corrected by enforcing partial decoding upon the interference decoder, even when complete decoding is possible. That is, we examine a variant of soft-IC-BP, where the interference decoder stops computing new messages (e.g. rightbound and leftbound) after the value $s_2^{(\ell)}$ (defined above) has dropped below a predetermined threshold. Once the optimization procedure has progressed and the design rate of the pair (λ, ρ) has sufficiently increased, it is possible to relax this enforcement.

C. Details of the Application of HK in Sec. V-B

Our discussion below assumes the results and notation of [25, Sec. III]. As noted in Sec. I-A, HK achieves its remarkable performance by constructing codes (here denoted $\mathcal{X}_1$ and $\mathcal{X}_2$) that are each obtained by combining two auxiliary codes, $\mathcal{U}_i$ and $\mathcal{W}_i$, $i = 1, 2$, with rates S_i and T_i, respectively. Destination 1, for example, decodes the codewords $\mathbf{u}_1 \in \mathcal{U}_1$ and $\mathbf{w}_1 \in \mathcal{W}_1$, produced at its corresponding source, as well as $\mathbf{w}_2 \in \mathcal{W}_2$, amounting to a partial decoding of $\mathcal{X}_2$.

The codes $\mathcal{U}_i$ and $\mathcal{W}_i$ are generated randomly, by independent selection of the components of their codewords according to the distributions of random variables U_i and W_i, respectively. In our application of HK, we assigned $U_i \sim$ Bernoulli(0.055) and $W_i \sim$ Bernoulli$(1/2)$. We also defined $X_i = \text{BPSK}(U_i \oplus W_i)$, $i = 1, 2$, where $\oplus$ denotes modulo-2 addition and the function BPSK maps the digits $\{0, 1\}$ to $\{1, -1\}$. This means that the codewords of $\mathcal{X}_i$ are similarly obtained by applying the above operation componentwise to

[28] To see this, observe that any constraint $|a| < b$ is equivalent to the two constraints $a < b$ and $a > -b$.

pairs of codewords $(\mathbf{u}_i, \mathbf{w}_i)$ from $\mathcal{U}_i$ and $\mathcal{W}_i$. An evaluation of the rate implied by [25, Theorem 3.1] gives 0.333 bits. More precisely, this figure is obtained by maximizing $S_1 + T_1 = S_2 + T_2$ (we restricted $S_1 = S_2$ and $T_1 = T_2$), as defined there, subject to [25, Equations (3.2)-(3.15)]. The maximizing choices were $S_i = 0.101$ bits and $T_i = 0.231$ bits, respectively.

Note that our use of binary rather than real-valued random variables as in applications of HK for the AWGN interference channel (e.g. [18]), as well as modulo-2 addition, follow from the channel's binary input alphabet.

Consider sequences of codes $\{\mathcal{X}_{i,n}\}_{n=1}^{\infty}$, $i = 1, 2$ where $\mathcal{X}_{i,n}$ has block length n, constructed as described above. We now verify that each such code-sequence is point-to-point "bad" for the BIAWGN channel, in the sense of Definition 4, with a probability that approaches 1 with n (the probability being derived from the above-mentioned random generation of the codes). Note that this assertion also holds by Theorem 7, relying on the fact that the rate of the code sequences (0.333) exceeds both R_{MUD} and R_{SUD} (see Sec. V-B). For simplicity of notation, we drop the indices i and n in the sequel.

To obtain our result, consider communication over a point-to-point BIAWGN channel using a code $\mathcal{X}$ generated as above. We wish to show that reliable communication requires an SNR that is greater than the Shannon limit for rate 0.333. Successful decoding in this setting recovers the codewords $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{W}$ which produced the transmitted $\mathbf{x} \in \mathcal{X}$, as byproducts. Thus, the communication setting is equivalent to that of a multiple-access channel (see e.g. [13, Sec. 14.3]), where two users transmit the codewords $\mathbf{u}$ and $\mathbf{w}$, respectively, and the receiver obtains,

$$\mathbf{y} = \text{BPSK}(\mathbf{u} \oplus \mathbf{w}) + \mathbf{z} \tag{100}$$

The operation BPSK is applied componentwise and $\mathbf{z}$ is zero-mean i.i.d AWGN noise whose components have variance $\sigma^2 = 1/\text{SNR}$. Invoking [13, Equations (14.99),(14.111)], the following conditions are necessary for reliable communications,

$$S \leq \frac{1}{n} I(\mathbf{U}; \mathbf{Y} \mid \mathbf{W}) + o(1) \tag{101}$$

$$T \leq \frac{1}{n} I(\mathbf{W}; \mathbf{Y} \mid \mathbf{U}) + o(1) \tag{102}$$

$$S + T \leq \frac{1}{n} I(\mathbf{U}, \mathbf{W}; \mathbf{Y}) + o(1) \tag{103}$$

Where the random vectors $\mathbf{U}$, $\mathbf{W}$ and $\mathbf{Y}$ correspond to the transmitted codewords from $\mathcal{U}$ and $\mathcal{W}$ and to the channel output, respectively. By methods similar to [63, Lemma 8], relying on the above-mentioned random construction of $\mathcal{U}$ and $\mathcal{W}$, and invoking the symmetry of the BIAWGN channel, we obtain, with a probability that approaches 1 with n,

$$\frac{1}{n} I(\mathbf{U}; \mathbf{Y} \mid \mathbf{W}) \leq I(U; Y \mid W) + o(1)$$

where U and W are independently distributed, and Y is related to them in the same way as (100). Thus, (101) translates to,

$$S \leq I(U; Y \mid W) + o(1) \tag{104}$$

Similarly, (102) and (103) imply,

$$T \leq I(W; Y \mid U) + o(1) \tag{105}$$

$$S + T \leq I(U, W; Y) + o(1) \tag{106}$$

An evaluation of (104), (105) and (106) reveals that the inequalities require an SNR of at least 0.7684 to be satisfied. This value, which is the minimum SNR required for reliable communications to be possible, exceeds the above-mentioned Shannon limit for rate 0.333 bits (Definition 1), which is $\text{SNR}^\star = 0.5941$. By Definition 4, this produces our desired result.

ACKNOWLEDGEMENTS

Suggestions by Rudiger Urbanke regarding the design of edge distributions for LDPC codes, and a discussion with Sergio Verdú, are gratefully acknowledged. Remarks by David Burshtein on a slide presentation of the material, and by Yuval Kochman on an initial draft of the paper, are very much appreciated.

REFERENCES

[1] A. Amraoui , S. Dusad and R. Urbanke, "Achieving General Points in the 2-User Gaussian MAC Without Time-sharing or Rate-Splitting by Means of Iterative Coding," in *International Symposium on Information Theory (ISIT2002)*, Lausanne, Switzerland, June 30–Jul. 5 2002.

[2] A. Ashikhmin, G. Kramer and S. ten Brink, "Extrinsic Information Transfer Functions: Model and Erasure Channel Properties," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2657–2673, 2004.

[3] O. Barak, U. Erez and D. Burshtein, "Bounds on Rates of LDPC Codes for BEC with Varying Erasure Rate," in *International Symposium on Information Theory (ISIT2008)*, Toronto, Ontario, Canada, July 6–11 2008.

[4] A. Bennatan, A. R. Calderbank, and S. Shamai (Shitz), "Bounds on the MMSE of "Bad" LDPC Codes at Rates Above Capacity," in *Forty-Sixth Annual Allerton Conference*, Illinois, Sep. 2008.

[5] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: turbo codes," in *Proceedings IEEE International Conference on Communications*, Geneva, Switzerland, 1993, pp. 1064–1070.

[6] J. Boutros and G. Caire, "Iterative Multiuser Joint Decoding: United Framework and Asymptotic Analysis," *IEEE Trans. on Inform. Theory*, vol. 48, no. 7, July 2002.

[7] D. Burshtein and G. Miller, "Asymptotic Enumeration Methods for Analyzing LDPC Codes," *IEEE Trans. Inf. Theory*, vol. 50, pp. 1115–1131, Jun. 2004.

[8] ——, "Expander Graph Arguments for Message-Passing Algorithms," *IEEE Trans. Inf. Theory*, vol. 47, pp. 782–790, Feb. 2004.

[9] D. Burshtein, M. Krivelevich, S. Litsyn, and G. Miller, "Upper Bounds on the Rate of LDPC Codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2437–2449, Sep. 2002.

[10] G. Caire, R. R. Muller and T. Tanaka, "Iterative Multiuser Joint Decoding: Optimal Power Allocation and Low-Complexity Implementation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1950–1973, Sep. 2004.

[11] M. H. M. Costa, A. A. El Gamal, "The Capacity Region of the Discrete Memoryless Interference Channel with Strong Interference (Corresp.)," *IEEE Trans. Info. Theory*, vol. 3, pp. 710–711, Sep. 1987.

[12] T. M. Cover and A. A. El Gamal, "Capacity Theorems for the Relay Channel," *IEEE Trans. Inf. Theory*, vol. 25, pp. 572–584, Sep. 1979.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[14] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson and R. L. Urbanke, "Finite-Length Analysis of Low-Density Parity-Check Codes on the Binary Erasure Channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1570–1579, June 2002.

[15] D. Divsalar, M. K. Simon, and D. Raphaeli, "Improved Parallel Interference Cancellation for CDMA," *IEEE Trans. Comm.*, vol. 46, pp. 258–268, Feb. 1998.

[16] P. Elias, "Coding for Noisy Channels," Mar. 1955, vol. 3, pp. 37–46.

[17] U. Erez, S. Shamai and R. Zamir, "Capacity and Lattice-Strategies for Cancelling Known Interference," *IEEE Trans. on Inform. Theory*, Nov. 2005.

[18] R. Etkin, D. Tse and H. Wang, "Gaussian Interference Channel Capacity to Within One Bit," *IEEE Trans. on Inform. Theory*, Dec. 2008.

[19] G. D. Forney, Jr., "The Viterbi Algorithm," in *Proc. IEEE*, vol. 61, Mar. 1973, pp. 268–276.

[20] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, Massachusetts: M.I.T Press, 1963.

[21] ——, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[22] K. S. Gomadam and S. A. Jafar, "Optimizing soft information in relay networks," in *Asilomar Conference on Signals, Systems and Computers*, 2006.

[23] D. Guo, S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

[24] S. H. Han and J. H. Lee, "Multi-Stage Partial Parallel Interference Cancellation Receivers for Multi-Rate DS-CDMA System," *IEICE Trans. Comm.*, vol. E86-B, pp. 170–180, Jan. 2003.

[25] T. S. Han and K. Kobayashi, "A New Achievable Rate Region for the Interference Channel," *IEEE Trans. Info. Theory*, vol. 27, pp. 49–60, Jan. 1981.

[26] A. Høst-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2020–2040, June 2005.

[27] Y.-H. Kim, "Coding Techniques for Primitive Relay Channels," in *Forty-Fifth Annual Allerton Conference*, Illinois, Sep. 2007, pp. 26–28.

[28] J. Kliewer, A. Mertins, "Soft-input Reconstruction of Binary Transmitted Quantized Overcomplete Expansions," *IEEE Signal Processing Letters*, vol. 11, pp. 899–903, Nov. 2004.

[29] M. Kobayashi, J. Boutros, and G. Caire, "Successive Interference Cancellation With SISO Decoding and EM Channel Estimation," *IEEE J. Sel. Areas in Comm.*, vol. 19, pp. 1450–1460, Aug. 2001.

[30] G. Kramer, "Cooperative Communications and Coding," in *Int. Symp. on Turbo Codes & Related Topics (talk)*, Lausanne, Switzerland, Sep. 2008.

[31] G. Kramer, I. Maric, and R. D. Yates, "Cooperative Communications," *Foundations and Trends in Networking*.

[32] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037–3063, Sep. 2005.

[33] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[34] P. Kumpoopong S. Kunaruttanapruk S. Jitapunkul, "Iterative Partial Soft Interference Cancellation with MMSE Multiuser Detection for Uplink Turbo-Coded MC-CDMA System," in *CNSR '05: Proceedings of the 3rd Annual Communication Networks and Services Research Conference*, Washington, DC, USA, 2005, pp. 218–222.

[35] C. Li, G. Yue, M. A. Khojastepour, X. Wang and M. Moadihian, "LDPC-Coded Cooperative Relay Systems: Performance Analysis and Code Design," *IEEE Trans. on Comm.*, vol. 46, pp. 485–495, Mar. 2008.

[36] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Efficient Erasure Correcting Codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569–584, Feb. 2001.

[37] ——, "Improved Low-Density Parity-Check Codes Using Irregular Graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, Feb. 2001.

[38] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[39] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "The Generalized Area Theorem and Some of Its Consequences," *IEEE Trans. Inform. Theory*, vol. 55, pp. 4793–4821, Nov. 2009.

[40] K. Narayanan, M. P. Wilson and A. Sprintson, "Joint Physical Layer Coding and Network Coding for Bi-Directional Relaying," in *Forty-Fifth Annual Allerton Conference*, Illinois, Sep. 2007.

[41] A. 'Nayagam and J. M. Shea and T. F. Wong, "Collaborative decoding of a broadcast message in bandwidth-constrained environments ," *IEEE J. Select Areas Commun.*, vol. 25, pp. 434–446, Feb. 2007.

[42] B. Nazer and M. Gastpar, "Compute-and-Forward: Harnessing Interference with Structured Codes," in *International Symposium on Information Theory (ISIT2008)*, Toronto, Ontario, Canada, July 6–11 2008.

[43] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping Set Distribution of LDPC Code Ensembles," *IEEE Trans. Inform. Theory*, vol. 51, pp. 929–953, Mar. 2005.

[44] D. Palomar and S. Verdú, "Representation of Mutual Information Via Input Estimates," *IEEE Trans. Inf. Theory*, vol. 53, pp. 453–470, Feb. 2007.

[45] M. Peleg, A. Sanderovich and S. Shamai (Shitz), "On Extrinsic Information of Good Codes Operating Over Gaussian Channels," *European Trans. Telecommunications*, vol. 18, pp. 133–139, 2007.

[46] T. Philosof and R. Zamir, "The Rate Loss of Single-Letter Characterization: The "Dirty" Multiple Access Channel," *IEEE Trans. Inf. Theory (submitted)*, Mar. 2008.

[47] T. Richardson and R. Urbanke, "The Capacity of Low-density Parity Check Codes Under Message-passing Decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[48] ——, *Modern Coding Theory*. Cambridge University Press, 2008.

[49] T. Richardson, A. Shokrollahi and R. Urbanke, "Design of Capacity-approaching Irregular Low-density Parity-check Codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[50] B. Rimoldi and R. Urbanke, "A Rate-splitting Approach to the Gaussian Multiple-access Channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 364–375, Mar. 1996.

[51] A. Roumy and D. Declercq, "Characterization and Optimization of LDPC Codes for the 2-User Gaussian Multiple Access Channel," *EURASIP Journal on Wireless Communications and Networking*, May 2007.

[52] A. Sanderovich, M. Peleg and S. Shamai (Shitz), "LDPC Coded MIMO Multiple Access With Iterative Joint Decoding," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1437–1450, Apr. 2005.

[53] I. Sason and R. Urbanke, "Parity-check density versus performance of binary linear block codes over memoryless symmetric channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1611–1635, July 2003.

[54] S. Shamai (Shitz) and S. Verdú, "The Empirical Distribution of Good Codes," *IEEE Trans. on Inform. Theory*, vol. 43, pp. 836–846, May. 1997.

[55] C. E. Shannon, "Two-way Communication Channels," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, Berkeley, CA, 1961.

[56] H. H. Sneessens and L. Vandendorpe, "Soft Decoding and Forward Improves Cooperative Communications," in *IEE Conf. on 3G and Beyond*, Nov. 2005.

[57] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. 27, pp. 533–547, Sep. 1981.

[58] E. C. van der Meulen, "Transmission of Information in a T-terminal Discrete Memoryless Channel," Ph.D. dissertation, Univ. California, Berkeley, Jun. 1968.

[59] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. London, U.K.: McGraw-Hill, 1979.

[60] Y. Li, B. Vucetic, T. F. Wong and M. Dohler, "Distributed Turbo Coding With Soft Information Relaying in Multihop Relay Networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, pp. 2040–2060, Nov. 2006.

[61] X. Wang, H. V. Poor, "Iterative (turbo) Soft Interference Cancellation and Decoding for Coded CDMA," *IEEE Trans. Comm.*, 1999.

[62] K.-M. Wu and C.-L. Wang, "An iterative multiuser receiver using partial parallel interference cancellation for turbo-coded DS-CDMA systems," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM 01)*, San Antonio, Texas, USA, Nov. 2001, pp. 244–248.

[63] A. D. Wyner, "The Wire-Tap Channel," *Bell Sys. Tech. Journal*, vol. 54, pp. 1355–1387, 1975.

[64] A. D. Wyner and J. Ziv, "The Rate-distortion Function for Source Coding with Side Information At the Decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, 1976.

[65] S. Yang and R. Koetter, "Network Coding Over a Noisy Relay: a Belief Propagation Approach," in *IEEE International Symposium on Information Theory (ISIT2007)*, Nice, France, 2007.